

PATENT APPLICATION

**Combinations of Genes for Producing Seed Plants Exhibiting Modulated
Reproductive Development**

Inventor(s):

Martin F. Yanofsky, a citizen of the United States, residing at:
5039 Manor Ridge Lane
San Diego, CA 92130

Soraya Pelaz, a citizen of the Spain, residing at:
C/ Gasómetro, 12 Portal 4 esc. drcha 2ºA
28005 Madrid, Spain

Gary Ditta, a citizen of the United States, residing at:
13815 Savage Way
Poway, CA 92064

Assignee:

University of California, San Diego
Technology Transfer and Intellectual Property Services
9500 Gilman Drive
La Jolla, CA 92093-0116

Entity:

Combinations of Genes for Producing Seed Plants Exhibiting Modulated Reproductive Development

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This work was supported by grant DCB-9018749 awarded by the National Science Foundation and by grant USDA 93-37304 awarded by the United States Department of Agriculture. The United States Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

A flower is the reproductive structure of a flowering plant. Following fertilization, the ovary of the flower becomes a fruit and bears seeds. As a practical consequence, production of fruit and seed-derived crops such as grapes, beans, corn, wheat, rice and hops is dependent upon flowering.

Early in the life cycle of a flowering plant, vegetative growth occurs, and roots, stems and leaves are formed. During the later period of reproductive growth, flowers as well as new shoots or branches develop. However, the factors responsible for the transition from vegetative to reproductive growth, and the onset of flowering, are poorly understood.

A variety of external signals, such as length of daylight and temperature, affect the time of flowering. The time of flowering also is subject to genetic controls that prevent young plants from flowering prematurely. Thus, the pattern of genes expressed in a plant is an important determinant of the time of flowering.

Given these external signals and genetic controls, a relatively fixed period of vegetative growth precedes flowering in a particular plant species. The length of time required for a crop to mature to flowering limits the geographic location in which it can be grown and can be an important determinant of yield. In addition, since the time of flowering determines when a plant is reproductively mature, the pace of a plant breeding program also depends upon the length of time required for a plant to flower.

Traditionally, plant breeding involves generating hybrids of existing plants, which are examined for improved yield or quality. The improvement of existing plant crops through plant breeding is central to increasing the amount of food grown in the world since

the amount of land suitable for agriculture is limited. For example, the development of new strains of wheat, corn and rice through plant breeding has increased the yield of these crops grown in underdeveloped countries such as Mexico, India and Pakistan. Unfortunately, plant breeding is inherently a slow process since plants must be reproductively mature before selective breeding can proceed.

For some plant species, the length of time needed to mature to flowering is so long that selective breeding, which requires several rounds of backcrossing progeny plants with their parents, is impractical. For example, perennial trees such as walnut, hickory, oak, maple and cherry do not flower for several years after planting. As a result, breeding of such plant species for insect or disease-resistance or to produce improved wood or fruit, for example, would require decades, even if only a few rounds of selection were performed.

Methods of promoting early reproductive development can make breeding of long generation seed plants such as trees practical for the first time. Methods of promoting early reproductive development also would be useful for shortening growth periods, thereby broadening the geographic range in which a crop such as rice, corn or coffee can be grown. Unfortunately, methods for promoting early reproductive development in a seed plant have not yet been described. Thus, there is a need for methods that promote early reproductive development. The present invention satisfies this need and provides related advantages as well.

DEFINITIONS

As used herein, the term "transgenic" refers to a seed plant that contains in its genome an exogenous nucleic acid molecule, which can be derived from the same or a different plant species. The exogenous nucleic acid molecule can be a gene regulatory element such as a promoter, enhancer or other regulatory element or can contain a coding sequence, which can be linked to a heterologous gene regulatory element.

As used herein, the term "seed plant" means an angiosperm or a gymnosperm. The term "angiosperm," as used herein, means a seed-bearing plant whose seeds are borne in a mature ovary (fruit). An angiosperm commonly is recognized as a flowering plant. The term "gymnosperm," as used herein, means a seed-bearing plant with seeds not enclosed in an ovary.

Angiosperms are divided into two broad classes based on the number of cotyledons, which are seed leaves that generally store or absorb food. Thus, a monocotyledonous angiosperm is an angiosperm having a single cotyledon, and a

dicotyledonous angiosperm is an angiosperm having two cotyledons. Angiosperms are well known and produce a variety of useful products including materials such as lumber, rubber, and paper; fibers such as cotton and linen; herbs and medicines such as quinine and vinblastine; ornamental flowers such as roses and orchids; and foodstuffs such as grains, oils, fruits and vegetables.

Angiosperms encompass a variety of flowering plants, including, for example, cereal plants, leguminous plants, oilseed plants, hardwood trees, fruit-bearing plants and ornamental flowers, which general classes are not necessarily exclusive. Such angiosperms include for example, a cereal plant, which produces an edible grain cereal. Such cereal plants include, for example, corn, rice, wheat, barley, oat, rye, orchardgrass, guinea grass, sorghum and turfgrass. In addition, a leguminous plant is an angiosperm that is a member of the pea family (*Fabaceae*) and produces a characteristic fruit known as a legume. Examples of leguminous plants include, for example, soybean, pea, chickpea, moth bean, broad bean, kidney bean, lima bean, lentil, cowpea, dry bean, and peanut. Examples of legumes also include alfalfa, birdsfoot trefoil, clover and sainfoin. An oilseed plant also is an angiosperm with seeds that are useful as a source of oil. Examples of oilseed plants include soybean, sunflower, rapeseed and cottonseed.

An angiosperm also can be a hardwood tree, which is a perennial woody plant that generally has a single stem (trunk). Examples of such trees include alder, ash, aspen, basswood (linden), beech, birch, cherry, cottonwood, elm, eucalyptus, hickory, locust, maple, oak, persimmon, poplar, sycamore, walnut and willow. Trees are useful, for example, as a source of pulp, paper, structural material and fuel.

An angiosperm also can be a fruit-bearing plant, which produces a mature, ripened ovary (usually containing seeds) that is suitable for human or animal consumption. For example, hops are a member of the mulberry family prized for their flavoring in malt liquor. Fruit-bearing angiosperms also include grape, orange, lemon, grapefruit, avocado, date, peach, cherry, olive, plum, coconut, apple and pear trees and blackberry, blueberry, raspberry, strawberry, pineapple, tomato, cucumber and eggplant plants. An ornamental flower is an angiosperm cultivated for its decorative flower. Examples of commercially important ornamental flowers include rose, orchid, lily, tulip and chrysanthemum, snapdragon, camellia, carnation and petunia plants. The skilled artisan will recognize that the methods of the invention can be practiced using these or other angiosperms, as desired.

Gymnosperms encompass four divisions: cycads, ginkgo, conifers and gnetophytes. The conifers are the most widespread of living gymnosperms and frequently are

cultivated for structural wood or for pulp or paper. Conifers include redwood trees, pines, firs, spruces, hemlocks, Douglas firs, cypresses, junipers and yews. The skilled artisan will recognize that the methods of the invention can be practiced with these and other gymnosperms.

5 As used herein, the term "non-naturally occurring seed plant" means a seed plant containing a genome that has been modified by man. A transgenic seed plant, for example, is a non-naturally occurring seed plant that contains an exogenous nucleic acid molecule and, therefore, has a genome that has been modified by man. Furthermore, a seed plant that contains, for example, a mutation in an endogenous floral meristem identity gene
10 regulatory element as a result of calculated exposure to a mutagenic agent also contains a genome that has been modified by man. In contrast, a seed plant containing a spontaneous or naturally occurring mutation is not a "non-naturally occurring seed plant" and, therefore, is not encompassed within the invention.

"Reproductive development" refers to the production of floral organs,
15 including but not limited to sepals, petal, stamens, carpels as well as pollen, ovules and/or seed. "Reproductive development" initiates upon the development of the floral meristem, typically derived from a shoot meristem.

The term "recombinant nucleic acid molecule," as used herein, means a
20 non-naturally occurring nucleic acid molecule that has been manipulated *in vitro* such that it is genetically distinguishable from a naturally occurring nucleic acid molecule. A recombinant nucleic acid molecule of the invention comprises two nucleic acid molecules that have been manipulated *in vitro* such that the two nucleic acid molecules are operably linked.

As used herein, the term "inducible regulatory element" means a nucleic acid
25 molecule that confers conditional expression upon an operably linked nucleic acid molecule, where expression of the operably linked nucleic acid molecule is increased in the presence of a particular inducing agent as compared to expression of the nucleic acid molecule in the absence of the inducing agent. In a method of the invention, a useful inducible regulatory element has the following characteristics: confers low level expression upon an operably
30 linked nucleic acid molecule in the absence of an inducing agent; confers high level expression upon an operably linked nucleic acid molecule in the presence of an appropriate inducing agent; and utilizes an inducing agent that does not interfere substantially with the normal physiology of a transgenic seed plant treated with the inducing agent. It is recognized, for example, that, subsequent to introduction into a seed plant, a particularly

useful inducible regulatory element is one that confers an extremely low level of expression upon an operably linked nucleic acid molecule in the absence of inducing agent. Such an inducible regulatory element is considered to be tightly regulated.

The term "operably linked," as used in reference to a regulatory element, such as a promoter or inducible regulatory element, and a nucleic acid molecule encoding a floral meristem identity gene product, means that the regulatory element confers regulated expression upon the operably linked nucleic acid molecule encoding the floral meristem identity gene product. Thus, the term operably linked, as used herein in reference to an inducible regulatory element and a nucleic acid molecule encoding a floral meristem identity gene product, means that the inducible regulatory element is linked to the nucleic acid molecule encoding a floral meristem identity gene product such that the inducible regulatory element increases expression of the floral meristem identity gene product in the presence of the appropriate inducing agent. It is recognized that two nucleic acid molecules that are operably linked contain, at a minimum, all elements essential for transcription, including, for example, a TATA box. One skilled in the art knows, for example, that an inducible regulatory element that lacks minimal promoter elements can be combined with a nucleic acid molecule having minimal promoter elements and a nucleic acid molecule encoding a floral meristem identity gene product such that expression of the floral meristem identity gene product can be increased in the presence of the appropriate inducing agent.

As used herein in reference to a nucleic acid molecule of the invention, the terms "sense" and "antisense" have their commonly understood meanings.

As used herein in reference to a nucleic acid molecule of the invention, the term "fragment" means a portion of the nucleic acid sequence containing at least about 50 base pairs to the full-length of the nucleic acid molecule. In contrast to an active fragment, as defined herein, a fragment of a nucleic acid molecule need not encode a functional portion of a gene product.

The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA

element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of flowering plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant, or a predecessor generation of the plant, by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include Agrobacterium-mediated transformation, biolistic methods, electroporation, in planta techniques, and the like.

The phrase "host cell" refers to a cell from any organism. Preferred host cells are derived from plants, bacteria, yeast, fungi, insects or other animals. Methods for introducing polynucleotide sequences into various types of host cells are well known in the art.

The "biological activity of a polypeptide" refers to any molecular activity or phenotype that is caused by the polypeptide. For example, the ability to transfer a phosphate to a substrate or the ability to bind a specific DNA sequence is a biological activity. One

biological activity of the gene products of the invention is the ability to modulate the time of development of reproductive structures in plants.

An "expression cassette" refers to a nucleic acid construct, which when introduced into a host cell, results in transcription and/or translation of an RNA or polypeptide, respectively. Antisense or sense constructs that are not or cannot be translated are expressly included by this definition.

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by reference to a specific nucleic acid sequence.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "nucleic acid encoding a gene product". In addition, the term specifically includes those sequences substantially identical (determined as described below) with an polynucleotide sequence disclosed here and that encode polypeptides that are either mutants of wild type polypeptides or retain the function of the polypeptide (e.g., resulting from conservative substitutions of amino acids in the polypeptides). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions,

the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, e.g., the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to a sequence or subsequence that has at least 40% sequence identity with a reference sequence. Alternatively, percent identity can be any integer from 40% to 100%. More preferred embodiments include at least: 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%. compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below. This definition also refers to the complement of a test sequence, when the test sequence has substantial identity to a reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP,

BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendrogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. For example, a reference sequence can be compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps.

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments;

or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid.

Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
 - 2) Aspartic acid (D), Glutamic acid (E);
 - 3) Asparagine (N), Glutamine (Q);
 - 4) Arginine (R), Lysine (K);
 - 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
 - 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (see, e.g., Creighton, *Proteins* (1984)).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30°C below the T_m . The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of

the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

In the present invention, genomic DNA or cDNA comprising nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a non-naturally occurring seed plant, the plant comprising: (1) a first ectopically expressed polynucleotide encoding an APETALA1 gene product at least 50% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8 or a CAULIFLOWER gene product at least 50% identical to SEQ ID NO:10 or SEQ ID NO:12; and (2) a second ectopically expressed nucleic acid molecule encoding a SEP1

gene product at least 50% identical to SEQ ID NO:28, a SEP2 gene product at least 50% identical to SEQ ID NO:30, a SEP3 gene product at least 50% identical to SEQ ID NO:32 or an AGL24 gene product at least 50% identical to SEQ ID NO:38. in some embodiments, the non-naturally occurring seed plant is characterized by early reproductive development. In some embodiments, expression of the first ectopically expressed polynucleotide is increased in a tissue of a plant compared to a wild type plant. In some embodiments, expression of the second ectopically expressed polynucleotide is increased in a tissue of a plant compared to a wild type plant. In some embodiments, expression of the first ectopically expressed polynucleotide is decreased in a tissue of a plant compared to a wild type plant. In some aspects, expression of the second ectopically expressed polynucleotide is decreased in a tissue of a plant compared to a wild type plant.

The invention provides for an endogenous first ectopically expressed polynucleotide comprising a modified gene regulatory element. Alternatively, the invention provides for an endogenous second ectopically expressed polynucleotide comprising a modified gene regulatory element. For example, the non-naturally occurring seed plant is a transgenic plant comprising a first exogenous gene regulatory element operably linked to the first ectopically expressible polynucleotide and a second exogenous gene regulatory element operably linked to the second ectopically expressible polynucleotide. In some aspects, the first polynucleotide is operably linked to the first exogenous gene regulatory element in a sense orientation. In some aspects, the first polynucleotide is operably linked to the first exogenous gene regulatory element in an antisense orientation. In some aspects, the second polynucleotide is operably linked to the second exogenous gene regulatory element in a sense orientation. In some aspects, the second polynucleotide is operably linked to the second exogenous gene regulatory element in an antisense orientation.

The invention also provides methods of modulating the timing of reproductive development in a plant, the methods comprising ectopically expressing a first polynucleotide encoding an APETALA1 gene product at least 50% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8 or a CAULIFLOWER gene product at least 50% identical to SEQ ID NO:10 or SEQ ID NO:12; and ectopically expressing a second nucleic acid molecule encoding a SEP1 gene product at least 50% identical to SEQ ID NO:28, a SEP2 gene product at least 50% identical to SEQ ID NO:30, a SEP3 gene product at least 50% identical to SEQ ID NO:32 or an AGL24 gene product at least 50% identical to SEQ ID NO:38. For example, in one aspect, the invention provides for introducing a first ectopically expressed nucleic acid molecule comprising a first polynucleotide encoding an APETALA1

gene product at least 50% identical to SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8 or a CAULIFLOWER gene product at least 50% identical to SEQ ID NO:10 or SEQ ID NO:12; and introducing a second ectopically expressed nucleic acid molecule comprising a second polynucleotide encoding a SEP1 gene product at least 50% identical to SEQ ID NO:28, a SEP2 gene product at least 50% identical to SEQ ID NO:30, a SEP3 gene product at least 50% identical to SEQ ID NO:32 or an AGL24 gene product at least 50% identical to SEQ ID NO:38.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides the surprising finding that ectopic expression of certain MADS-box-containing gene products, such as SEP1, SEP2, SEP3 or AGL24, combined with the ectopic expression of AP1, CAL or LFY gene products, result in modulated reproductive development. Thus, this invention provides plants comprising such ectopically expressible gene products as well as methods of modulating the timing of reproductive development in plants.

A flower, like a leaf or shoot, is derived from the shoot apical meristem, which is a collection of undifferentiated cells set aside during embryogenesis. The production of vegetative structures, such as leaves or shoots, and of reproductive structures, such as flowers, is temporally segregated, such that a leaf or shoot arises early in a plant life cycle, while a flower develops later. The transition from vegetative to reproductive development is the consequence of a process termed floral induction (Yanofsky, *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 46:167-188 (1995), which is incorporated herein by reference).

Once induced, shoot apical meristem either persists and produces floral meristem, which gives rise to flowers, and lateral meristem, which gives rise to branches, or is itself converted to floral meristem. Floral meristem differentiates into a single flower having a fixed number of floral organs in a whorled arrangement. Dicots, for example, contain four whorls (concentric rings), in which sepals (first whorl) and petals (second whorl) surround stamens (third whorl) and carpels (fourth whorl).

Following the transition from vegetative to reproductive development in *Arabidopsis*, flower meristems arise on the flanks of the shoot apical (inflorescence) meristem and subsequently develop into flowers with four organ types (sepals, petals, stamens and carpels). Flower meristem identity is specified in part by the *APETALA1* (*AP1*), *CALIFLOWER* (*CAL*) and *LEAFY* (*LFY*) genes. In *ap1* mutants, the sepals are transformed to leaf-like organs and the petals fail to develop. In the axils of these leaf-like organs,

secondary flowers arise which repeat the same pattern as the primary ones. Although *cal* single mutants appear wild type, *apl cal* double mutants display a massive proliferation of inflorescence-like meristems in positions that would normally be occupied by solitary flowers. The functional redundancy shared by *API* and *CAL* can be explained in part by the fact that these two genes encode related members of the MADS box family of regulatory proteins (Bowman *et al.*, *Development* 119, 721-743 (1993); Gustafson-Brown *et al.*, *Cell* 76, 131-143 (1994); Kempin *et al.*, *Science* 267, 522-525 (1995); Mandel *et al.*, *Nature* 360, 273-277 (1992)).

Genetic studies led to the proposal of the ABC model that explains how the individual and combined activities of the ABC genes specify the four organ types of the typical eudicot flower. A alone specifies sepals, A and B specify petals, B and C specify stamens, and C alone specifies carpels. In *Arabidopsis*, the A-function genes are *API* and *APETALA2* (*AP2*), B-function genes are *APETALA3* (*AP3*), *PISTILLATA* (*PI*), and the C-function gene is *AGAMOUS* (*AG*). In addition, recent studies have shown that a trio of closely related genes, *SEPALLATA1/2/3* (*SEP1/2/3*), are required for petal, stamen and carpel identity, and are thus necessary for the activities of the B- and C-function genes (Pelaz *et al.*, *Nature* 405, 200-203 (2000)). Remarkably, with the exception of the *AP2* gene, all of the other organ identity genes belong to the extended family of MADS-box genes, a family that is known to include more than 44 distinct sequences in *Arabidopsis* (Alvarez-Buylla *et al.*, *Proc. Natl. Acad. Sci. USA* 97, 5328-5333 (2000); Davies and Schwarz-Sommer, In *Plant Promoters and Transcription Factors (Results and Problems in Cell Differentiation 20)*, (ed. L. Nover), pp. 235-258 (1994); Purugganan *et al.*, *Genetics* 140, 345-356 (1995); Rounsley *et al.*, *Plant Cell* 7, 1259-1269 (1995)).

MADS-domain proteins, well characterized in yeast (MCM1, Ammerer, *Genes Dev.* 4, 299-312 (1990)) and mammals (SRF, Norman *et al.*, *Cell* 55, 989-1003 (1988)) form dimers that bind to DNA and form ternary complexes with many unrelated proteins (Lamb and McKnight, *Trends Biochem. Sci.* 16, 417-422 (1991); Shore and Sharrocks, *Eur. J. Chem.* 229, 1-13 (1995)). A number of studies have shown that heterodimers and ternary complexes of plant MADS-domain proteins can occur, and given the overlapping expression pattern of numerous MADS-box genes, such interactions greatly increase the regulatory complexity of MADS-box genes (Davies *et al.*, *EMBO J.* 15, 4330-4343 (1996); Egea-Cortines *et al.*, *EMBO J.* 18, 5370-5379 (1999); Fan *et al.*, *Plant J.* 11, 999-1010 (1997)). The regulatory specificity of these genes is achieved through protein-protein interactions and not through different intrinsic DNA binding specificities (Krizek and

Meyerowitz, *Proc. Natl. Acad. Sci. USA* 93, 4063-4070 (1996); Shore and Sharrocks, *Eur. J. Chem.* 229, 1-13 (1995)). MADS box proteins are composed of four different domains, designated M, I, K and C. The MADS (M) domain, is highly conserved among these proteins, and is responsible for the binding to DNA in addition to its participation in

5 homodimer formation of some proteins. The I region also participates in homodimer formation (Krizek and Meyerowitz, *supra*; Riechmann *et al.*, *Proc. Natl. Acad. Sci. USA* 93, 4793-4798 (1996)). Adjacent to the I region is the K-domain, so named, due to its similarity to the coiled-coil domain of keratin. It is absent in the non-plant proteins, and has been implicated in protein-protein interaction (Fan *et al.*, *supra*; Krizek and Meyerowitz, *supra*;

10 Mizukami *et al.*, *Plant Cell* 8, 831-845 (1996); Moon *et al.*, *Plant Physiol.* 120, 1193-1203 (1999); Riechmann *et al.*, *supra*). The C-terminal region has been proposed to be involved in transcriptional activation (Huang *et al.*, *Plant Mol. Biol.* 28, 549-567 (1995)), and also to play a role in the formation of ternary complexes (Egea-Cortines *et al.*, *EMBO J.* 18, 5370-5379 (1999)).

15 Although shoot meristem and floral meristem both consist of meristemic tissue, shoot meristem is distinguishable from the more specialized floral meristem. Shoot meristem generally is indeterminate and gives rise to an unspecified number of floral and lateral meristems. In contrast, floral meristem is determinate and gives rise to the fixed number of floral organs that comprise a flower.

20 By convention herein, a wild-type gene sequence is represented in upper case italic letters (for example, *APETALA1*), and a wild-type gene product is represented in upper case non-italic letters (APETALA1). Further, a mutant gene allele is represented in lower case italic letters (*ap1*), and a mutant gene product is represented in lower case non-italic letters (ap1).

25 Genetic studies have identified a number of genes involved in regulating flower development. These genes can be classified into different groups depending on their function. Flowering time genes, for example, are involved in floral induction and regulate the transition from vegetative to reproductive growth. In comparison, the floral meristem identity genes, which are the subject matter of the present invention as disclosed herein,

30 encode proteins that promote the conversion of shoot meristem to floral meristem in an angiosperm. In addition, floral organ identity genes encode proteins that determine whether sepals, petals, stamens or carpels are formed during floral development (Yanofsky, *supra*, 1995; Weigel, *Ann. Rev. Genetics* 29:19-39 (1995), which is incorporated herein by

reference). Some of the floral meristem identity gene products also have a role in specifying floral organ identity.

Floral meristem identity genes have been identified by characterizing genetic mutations that prevent or alter floral meristem formation. Among floral meristem identity gene mutations in *Arabidopsis thaliana*, those in the gene *LEAFY* (*LFY*) generally have the strongest effect on floral meristem identity. Mutations in *LFY* completely transform the basal-most flowers into secondary shoots and have variable effects on later-arising (apical) flowers. In comparison, mutations in the floral meristem identity gene *APETALA1* (*AP1*) result in replacement of a few basal flowers by inflorescence shoots that are not subtended by leaves. An apical flower produced in an *ap1* mutant has an indeterminate structure, in which a flower arises within a flower. These mutant phenotypes indicate that both *AP1* and *LFY* contribute to establishing the identity of the floral meristem although neither gene is absolutely required. The phenotype of *lfy ap1* double mutants, in which structures with flower-like characteristics are very rare, indicates that *LFY* and *AP1* encode partially redundant activities.

In addition to the *LFY* and *AP1* genes, a third locus that greatly enhances the *ap1* mutant phenotype has been identified in *Arabidopsis*. This locus, designated *CAULIFLOWER* (*CAL*), derives its name from the resulting "cauliflower" phenotype, which is strikingly similar to the common garden variety of cauliflower (Kempin et al., *Science* 267:522-525 (1995), which is incorporated herein by reference). In an *ap1 cal* double mutant, floral meristem behaves as shoot meristem in that there is a massive proliferation of meristems in the position that normally would be occupied by a single flower. However, an *Arabidopsis* mutant lacking only *CAL*, such as *cal-1*, has a normal phenotype, indicating that *AP1* can substitute for the loss of *CAL* in these plants. In addition, because floral meristem that forms in an *ap1* mutant behaves as shoot meristem in an *ap1 cal* double mutant, *CAL* can largely substitute for *AP1* in specifying floral meristem. These genetic data indicate that *CAL* and *AP1* encode activities that are partially redundant in converting shoot meristem to floral meristem.

Other genetic loci play at least minor roles in specifying floral meristem identity. For example, although a mutation in *APETALA2* (*AP2*) alone does not result in altered inflorescence characteristics, *ap2 ap1* double mutants have indeterminate flowers (flowers with shoot-like characteristics; Bowman et al., *Development* 119:721-743 (1993), which is incorporated herein by reference). Also, mutations in the *CLAVATA1* (*CLV1*) gene result in an enlarged meristem and lead to a variety of phenotypes (Clark et al., *Development*

119:397-418 (1993)). In a *clv1 apl* double mutant, formation of flowers is initiated, but the center of each flower often develops as an indeterminate inflorescence. Thus, mutations in *CLAVATA1* result in the loss of floral meristem identity in the center of wild-type flowers. Genetic evidence also indicates that the gene product of *UNUSUAL FLORAL ORGANS* (UFO) plays a role in determining the identity of floral meristem. Additional floral meristem identity genes associated with altered floral meristem formation remain to be isolated.

Mutations in another locus, designated *TERMINAL FLOWER (TFL)*, produce phenotypes that generally are reversed as compared to mutations in the floral meristem identity genes. For example, *tfl* mutants flower early, and the indeterminate apical and lateral meristems develop as determinate floral meristems (Alvarez et al., *Plant J.* 2:103-116 (1992)). These characteristics indicate that the TFL promotes maintenance of shoot meristem. *TFL* also acts directly or indirectly to negatively regulate AP1 and LFY expression in shoot meristem since these AP1 and LFY are ectopically expressed in the shoot meristem of *tfl* mutants (Gustafson-Brown et al., *Cell* 76:131-143 (1994); Weigel et al., *Cell* 69:843-859 (1992)). It is recognized that a plant having a mutation in TFL can have a phenotype similar to a non-naturally occurring seed plant of the invention. Such *tfl* mutants, however, which have a mutation in an endogenous *TERMINAL FLOWER* gene, are explicitly excluded from the scope of the present invention.

The results of such genetic studies indicate that several floral meristem identity gene products, including AP1, CAL and LFY, act redundantly to convert shoot meristem to floral meristem in an angiosperm. As disclosed herein, ectopic expression of a single floral meristem identity gene product such as AP1, CAL or LFY is sufficient to convert shoot meristem to floral meristem in an angiosperm. Thus, the present invention provides a non-naturally occurring seed plant such as an angiosperm or gymnosperm that contains a first or second ectopically expressible nucleic acid molecule encoding a first floral meristem identity gene product, provided that such ectopic expression is not due to a mutation in an endogenous *TERMINAL FLOWER* gene.

As disclosed herein, an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product can be, for example, a transgene encoding a floral meristem identity gene product under control of a heterologous gene regulatory element. In addition, such an ectopically expressible nucleic acid molecule can be an endogenous floral meristem identity gene coding sequence that is placed under control of a heterologous gene regulatory element. The ectopically expressible nucleic acid molecule also can be, for example, an endogenous floral meristem identity gene having a modified gene regulatory

element such that the endogenous floral meristem identity gene is no longer subject to negative regulation by TFL.

The term "ectopically expressible" is used herein to refer to a nucleic acid molecule encoding a floral meristem identity gene product that can be expressed in a tissue other than a tissue in which it normally is expressed or at a time other than the time at which it normally is expressed, provided that the floral meristem identity gene product is not expressed from its native, naturally occurring promoter. Ectopic expression of a floral meristem identity gene product is a result of the expression of the gene coding region from a heterologous promoter or from a modified variant of its own promoter, such that expression of the floral meristem identity gene product is no longer in the tissue in which it normally is expressed or at the time at which it normally is expressed. An exogenous nucleic acid molecule encoding an AP1 gene product under control of its native, wild type promoter, for example, does not constitute an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product. However, a nucleic acid molecule encoding an AP1 gene product under control of a constitutive promoter, which results in expression of AP1 in a tissue such as shoot meristem where it is not normally expressed, is an ectopically expressible nucleic acid molecule as defined herein.

Actual ectopic expression of a floral meristem identity gene is dependent on various factors and can be constitutive or inducible expression. For example, AP1, which normally is expressed in floral meristem, is ectopically expressible in the shoot meristem of an angiosperm. When a floral meristem identity gene product such as AP1, CAL or LFY is ectopically expressed in shoot meristem in an angiosperm, the shoot meristem is converted to floral meristem and early reproductive development can occur (*see* WO 97/46078, incorporated herein by reference).

An ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product can be expressed prior to the time in development at which the corresponding endogenous gene normally is expressed. For example, an *Arabidopsis* plant grown under continuous light conditions expresses *AP1* just prior to day 18, when normal reproductive development (flowering) begins. However, *AP1* can be ectopically expressed in shoot meristem prior to day 18, resulting in early conversion of shoot meristem to floral meristem and early reproductive development. *See* WO 97/46078. As disclosed in Example ID of WO 97/46078, a transgenic *Arabidopsis* plant that ectopically expresses *AP1* in shoot meristem under control of a constitutive promoter can flower at day 10, which is earlier than the time of reproductive development for a non-transgenic plant grown under the same

conditions (day 18). It is recognized that in some transgenic seed plants containing, for example, an exogenous nucleic acid molecule encoding AP1 under control of a constitutive promoter, neither the exogenous nor endogenous AP1 will be expressed. Such transgenic plants in which AP1 gene expression is cosuppressed, although not characterized by early reproductive development, also can be valuable as disclosed below.

I. Floral Meristem Gene Products

As used herein, the term "floral meristem identity gene product" means a gene product that promotes conversion of shoot meristem to floral meristem in an angiosperm.

Expression of a floral meristem identity gene product such as AP1, CAL or LFY in shoot meristem can convert shoot meristem to floral meristem in an angiosperm. Furthermore, ectopic expression of a floral meristem identity gene product also can promote early reproductive development.

A floral meristem identity gene product is distinguishable from a late flowering gene product or an early flowering gene product. The use of a late flowering gene product or an early flowering gene product is not encompassed within the scope of the present invention. In addition, reference is made herein to an "inactive" floral meristem identity gene product, as exemplified by the product of the *Brassica oleracea* var. *botrytis* CAL gene (*BobCAL*) (see below). Expression of an inactive floral meristem identity gene product in an angiosperm does not result in the conversion of shoot meristem to floral meristem in the angiosperm. An inactive floral meristem identity gene product such as BobCAL is excluded from the meaning of the term "floral meristem identity gene product" as defined herein.

A. AP1

A floral meristem identity gene product can be, for example, an AP1 gene product having the amino acid sequence of SEQ ID NO: 2, which is a 256 amino acid gene product encoded by the *Arabidopsis thaliana* AP1 cDNA. The *Arabidopsis* AP1 cDNA encodes a highly conserved MADS domain, which can function as a DNA-binding domain, and a K domain, which has structural similarity to the coiled-coil domain of keratins and can be involved in protein-protein interactions.

As used herein, the term "APETALA1," "AP1" or "AP1 gene product" means a floral meristem identity gene product that is characterized, in part, by having an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO: 2 in the region from amino acid 1 to amino acid 163 or with the amino acid sequence of SEQ ID NO: 8 in the region from amino acid 1 to amino acid 163. Alternatively, "AP1 gene product" refers to a gene product substantially identical to SEQ ID NO:2 or SEQ ID NO:8. Like other floral meristem identity gene products, AP1 promotes conversion of shoot meristem to floral meristem in an angiosperm. An AP1 gene product useful in the invention can be, for example, *Arabidopsis* AP1 having the amino acid sequence of SEQ ID NO: 2; *Brassica oleracea* AP1 having the amino acid sequence of SEQ ID NO: 4; *Brassica oleracea* var. *botrytis* AP1 having the amino acid sequence of SEQ ID NO: 6 or *Zea mays* AP1 having the amino acid sequence of SEQ ID NO: 8.

In wild-type *Arabidopsis*, *AP1* RNA is expressed in flowers but is not detectable in roots, stems or leaves (Mandel et al., *Nature* 360:273-277 (1992), which is incorporated herein by reference). The earliest detectable expression of *AP1* RNA is in young floral meristem at the time it initially forms on the flanks of shoot meristem. Expression of *AP1* increases as the floral meristem increases in size; no *AP1* expression is detectable in shoot meristem. In later stages of development, *AP1* expression ceases in cells that will give rise to reproductive organs of a flower (stamens and carpels), but is maintained in cells that will give rise to non-reproductive organs (sepals and petals; Mandel, *supra*, 1992). Thus, in nature, *AP1* expression is restricted to floral meristem and to certain regions of the flowers that develop from this meristemic tissue.

B. CAL

CAULIFLOWER (CAL) is another example of a floral meristem identity gene product. As used herein, the term "CAULIFLOWER," "CAL" or "CAL gene product" means a floral meristem identity gene product that is characterized, in part, by substantial identity to an amino acid sequence of SEQ ID NO: 10 in the region from amino acid 1 to amino acid 160 or with the amino acid sequence of SEQ ID NO: 12 in the region from amino acid 1 to amino acid 160. Alternatively, "CAL gene product" refers to a gene product substantially identical to SEQ ID NO:10 or SEQ ID NO:12.

A CAL gene product is exemplified by the *Arabidopsis* CAL gene product, which has the amino acid sequence of SEQ ID NO: 10, or the *Brassica oleracea* CAL gene

product, which has the amino acid sequence of SEQ ID NO:12. As disclosed herein, CAL, like AP1, contains a MADS domain and a K domain. The MADS domains of CAL and AP1 differ in only five of 56 amino acid residues, where four of the five differences represent conservative amino acid replacements. Over the entire sequence, the *Arabidopsis* CAL and *Arabidopsis* AP1 sequences (SEQ ID NOS: 10 and 2) are 76% identical and are 88% similar if conservative amino acid substitutions are allowed.

Similar to the expression pattern of *AP1*, *CAL* RNA is expressed in young floral meristem in *Arabidopsis*. However, in contrast to *AP1* expression, which is high throughout sepal and petal development, *CAL* expression is low in these organs. Thus, in nature, *CAL* is expressed in floral meristem and, to a lesser extent, in the organs of developed flowers.

The skilled artisan will recognize that, due to the high sequence conservation between AP1 and CAL, a novel ortholog can be categorized as both a CAL and an AP1, as defined herein. However, if desired, an AP1 ortholog can be distinguished from a CAL ortholog by demonstrating a greater similarity to *Arabidopsis* AP1 than to any other MADS box gene, such as CAL, as set forth in Purugganan et al. (*Genetics* 140:345-356 (1995), which is incorporated herein by reference). Furthermore, AP1 can be distinguished from CAL by its ability to complement, or restore a wild-type phenotype, when introduced into a strong *ap1* mutant. For example, introduction of *Arabidopsis* AP1 (*AGL7*) complements the phenotype of the strong *ap1-1* mutant; however, introduction of *CAL* (*AGL10*) into a *cal-1 ap1-1* mutant plant yields the *ap1-1* single mutant phenotype, demonstrating that *CAL* cannot complement the *ap1-1* mutation (see, for example, Mandel et al., *supra*, 1992; Kempin et al., *supra*, 1995). Thus, AP1 can be distinguished from CAL, if desired, by the ability of a nucleic acid molecule encoding AP1 to complement a strong *ap1* mutant such as *ap1-1* or *ap1-15*.

C. LFY

LEAFY (LFY) is yet another example of a floral meristem identity gene product. As used herein, the term "LEAFY" or "LFY" or "LFY gene product" means a floral meristem identity gene product that is characterized, in part, by having an amino acid sequence that has substantial identity with the amino acid sequence of SEQ ID NO:16. In nature, LFY is expressed in floral meristem as well as during vegetative development. As disclosed herein, ectopic expression in shoot meristem of a floral meristem identity gene

product, which normally is expressed in floral meristem, can convert shoot meristem to floral meristem in an angiosperm. Under appropriate conditions, ectopic expression in shoot meristem of a floral meristem identity gene product such as AP1, CAL, LFY, or a combination thereof, can promote early reproductive development.

5

D. Floral Meristem Gene Product Orthologs

Flower development in *Arabidopsis* is recognized in the art as a model for flower development in angiosperms in general. Gene orthologs corresponding to the *Arabidopsis* genes involved in the early steps of flower formation have been identified in
10 distantly related angiosperm species, and these gene orthologs show remarkably similar patterns of RNA expression. Mutations in gene orthologs also result in phenotypes that correspond to the phenotype produced by a similar mutation in *Arabidopsis*. For example, orthologs of the *Arabidopsis* floral meristem identity genes *AP1* and *LFY* and the *Arabidopsis* organ identity genes *AGAMOUS*, *APETALA3* and *PISTILLATA* have been isolated from
15 monocots such as maize and, where characterized, reveal the anticipated RNA expression patterns and related mutant phenotypes (Schmidt et al., *Plant Cell* 5:729-737 (1993); and Veit et al., *Plant Cell* 5:1205-1215 (1993), each of which is incorporated herein by reference). Furthermore, a gene ortholog can be functionally interchangeable in that it can function across distantly related species boundaries (Mandel et al., *Cell* 71:133-143 (1992), which is
20 incorporated herein by reference). Taken together, these data suggest that the underlying mechanisms controlling the initiation and proper development of flowers are conserved across distantly related dicot and monocot boundaries.

Floral meristem identity genes in particular are conserved among distantly related angiosperm and gymnosperm species. For example, a gene ortholog of *Arabidopsis*
25 *AP1* has been isolated from *Antirrhinum majus* (snapdragon; Huijser et al., *EMBO J.* 11:1239-1249 (1992), which is incorporated herein by reference). An ortholog of *Arabidopsis AP1* also has been isolated from *Brassica oleracea* var. *botrytis* (cauliflower, see SEQ ID NO:6), *Zea Mays* (maize; see SEQ ID NO:8) and rice (OsMADS14 (*Plant Physiology* 120:1193-1203 (1999)). Furthermore, *AP1* orthologs also can be isolated from
30 gymnosperms. Similarly, gene orthologs of *Arabidopsis LFY* have been isolated from angiosperms such as *Antirrhinum majus*, tobacco and poplar tree and from gymnosperms such as Douglas fir (Coen et al., *Cell*, 63:1311-1322 (1990); Kelly et al., *Plant Cell* 7:225-234 (1995); and Rottmann et al., *Cell Biochem. Suppl.* 17B: 23 (1993); Strauss et al.,

Molec. Breed 1:5-26 (1995), each of which is incorporated herein by reference). The conservation of floral meristem identity gene products in non-flowering plants such as coniferous trees indicates that floral meristem identity genes can promote the reproductive development of gymnosperms as well as angiosperms.

5 The characterization of *ap1* and *lfy* mutants also indicates that floral meristem identity gene products such as AP1 and LFY function similarly in distantly related plant species. For example, a mutation in the *Antirrhinum* *API* ortholog results in a phenotype similar to the *Arabidopsis* *ap1* indeterminate flower within a flower phenotype (Huijser et al., *supra*, 1992). In addition, a mutation in the *Antirrhinum* *LFY* ortholog results in a phenotype
10 similar to the *Arabidopsis* *lfy* mutant phenotype (Coen et al., *supra*, 1995)

 A floral meristem identity gene product also can function across species boundaries. For example, introduction of a nucleic acid molecule encoding *Arabidopsis* LFY into a heterologous seed plant such as tobacco or aspen results in early reproductive development (Weigel and Nilsson, *Nature* 377:495-500 (1995), which is incorporated herein
15 by reference). As disclosed herein, a nucleic acid molecule encoding an *Arabidopsis* AP1 gene product (SEQ ID NO: 1) or an *Arabidopsis* CAL gene product (SEQ ID NO: 9) can be introduced into a heterologous seed plant such as corn, wheat, rice or pine and, upon ectopic expression, can promote early reproductive development in the transgenic seed plant. Furthermore, as disclosed herein, the conserved nature of the *API*, *CAL* and *LFY* coding
20 sequences among diverse seed plant species allows a nucleic acid molecule encoding a floral meristem identity gene product isolated from essentially any seed plant to be introduced into essentially any other seed plant, wherein, upon appropriate expression of the introduced nucleic acid molecule in the seed plant, the floral meristem identity gene product promotes early reproductive development in the seed plant.

25 If desired, a novel *API*, *CAL* or *LFY* coding sequence can be isolated from a seed plant using a nucleotide sequence as a probe and methods well known in the art of molecular biology (Sambrook et al. (eds.), *Molecular Cloning: A Laboratory Manual* (Second Edition), Plainview, NY: Cold Spring Harbor Laboratory Press (1989), which is incorporated herein by reference). As exemplified herein and discussed in detail below (see
30 Example VA), an *API* ortholog from *Zea Mays* (maize; SEQ ID NO: 7) was isolated using the *Arabidopsis* *API* cDNA (SEQ ID NO: 1) as a probe.

II. AGAMOUS-like Gene Products

Modulation of expression of the gene products described below, either alone, or in combination with the ectopic expression of AP1 or CAL, results in the modulation of the development of reproductive development in plants.

A. SEP1, SEP2 and SEP3

SEP1, SEP2 and SEP3 (previously known as AGL2, AGL4 and AGL9, respectively) are a class of floral organ identity gene products that are required for development of stamens and carpels (Pelaz, *et al.*, *Nature* 405:200-203 (2000)). The SEP gene products are functionally redundant. Therefore, inactivation of only one SEP gene product does not typically result in the development of a mutant floral phenotype. Ectopic or increased expression of a SEP gene product results in early development of reproductive structures. Delay of reproductive development typically requires the reduction of expression of at least two and sometimes all three SEP gene products due to the redundancy of their function.

As used herein, the term "SEP1" or "SEP1 gene product" means a floral meristem identity gene product, or active fragment thereof, that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 28. The term "SEP2" or "SEP2 gene product" means a floral meristem identity gene product, or active fragment thereof, that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 30. SEP1 and SEP2 sequences were previously described in Ma *et al.*, *Genes & Development* 5:484-495 (1991). An exemplary SEP1 nucleic acid sequence is displayed as SEQ ID NO:27. An exemplary SEP2 nucleic acid sequence is displayed as SEQ ID NO:29. The term "SEP3" or "SEP3 gene product" means a floral meristem identity gene product, or active fragment thereof, that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 32. SEP3 sequences were previously described in Mandel *et al.*, *Sex. Plant Reprod.* 11:22-28 (1998). An exemplary SEP3 nucleic acid sequence is displayed as SEQ ID NO:31.

B. AGL20

As used herein, the term "AGL20" or "AGL20 gene product" means a gene product that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 34. AGL20 is also known as "SOC1." See, e.g., Samach *et al.* *Science* 288:1613-1616 (2000). Reduction of endogenous expression of AGL20 results in delayed development of reproductive structures in plants. An exemplary AGL20 nucleic acid sequence is displayed as SEQ ID NO:33.

C. AGL22

As used herein, the term "AGL22" or "AGL22 gene product" means a gene product that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 36. Decreased expression of an AGL22 gene product results in early development of reproductive structures. AGL22 is also known as "SVP." An exemplary AGL22 nucleic acid sequence is displayed as SEQ ID NO:35.

D. AGL24

As used herein, the term "AGL24" or "AGL24 gene product" means a gene product that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 38. An exemplary AGL24 nucleic acid sequence is displayed as SEQ ID NO:37. Ectopic or increased expression of AGL24 results in early development of reproductive structures in plants. Reduced expression of endogenous AGL24 results in delayed development of reproductive structures in plants.

E. AGL27

As used herein, the term "AGL27" or "AGL27 gene product" means a gene product that is characterized, in part, by having an amino acid sequence substantially identical to SEQ ID NO: 36. An exemplary AGL27 cDNA nucleic acid sequence is displayed as SEQ ID NO:39. An alternatively spliced AGL27 cDNA, and resulting translated product, are displayed as SEQ ID NO:48 and SEQ ID NO:49.

III. Effect of Gene Products of the Invention on Timing of Reproductive Development

As described in U.S. Patent 6,002,069, ectopic expression of the AP1 or CAL gene products results in the early development of reproductive structures in plants. The present invention demonstrates that ectopic expression of a number of other genes in combination with the ectopic expression of AP1 or CAL, leads to significantly earlier timing of reproductive development than plants ectopically expressing AP1 or CAL alone. In one embodiment, the invention provides a non-naturally occurring seed plant that contains a first ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product, provided that the first nucleic acid molecule is not ectopically expressed due to a

mutation in an endogenous *TERMINAL FLOWER* gene, and a second ectopically expressible nucleic acid molecule encoding SEP1, SEP2, SEP3 or AGL24, wherein the plant is characterized by modulated timing of reproductive development.

As used herein, the term "characterized by early reproductive development," when used in reference to a non-naturally occurring seed plant of the invention, means a non-naturally occurring seed plant that forms reproductive structures at an earlier stage than when reproductive structures form on a corresponding naturally occurring seed plant that is grown under the same conditions and that does not ectopically express a floral meristem identity gene product. In addition, "characterized by early reproductive development" also refers to the formation of reproduction structures at an earlier stage than a plant identical except for the lack of ectopic expression of the nucleic acids of the invention (e.g., polynucleotides substantially similar to nucleic acid molecules encoding SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38 or SEQ ID NO:40). Note that "stage," as used above, refers to either the amount of time from germination of seed or the number of leaves that a plant produces prior to initiation of reproductive structures. Similarly, "characterized by late reproductive development" or "characterized by delayed reproductive development" refers to the delayed development of reproductive structures compared to a naturally-occurring seed plant or to a plant, natural or transgenic, that does not ectopically express a nucleic acid of the invention. The reproductive structure of an angiosperm, for example, is a flower, and the reproductive structure of a coniferous plant is a cone. For a particular naturally occurring seed plant, reproductive development occurs at a well-defined time that depends, in part, on genetic factors as well as on environmental conditions, such as day length and temperature. Thus, given a defined set of environmental condition and lacking ectopic expression of a floral meristem identity gene product, a naturally occurring seed plant will undergo reproductive development at a relatively fixed time.

It is recognized that various transgenic plants that are characterized by altered timing of reproductive development have been described previously. Such transgenic plants, as discussed herein, are distinguishable from a non-naturally occurring seed plant of the invention or are explicitly excluded from the present invention. The product of a "late-flowering gene" can promote early reproductive development. However, a late flowering gene product is not a floral meristem identity gene product since it does not specify the conversion of shoot meristem to floral meristem in an angiosperm. Therefore, a

transgenic plant expressing a late-flowering gene product is distinguishable from a non-naturally occurring seed plant of the invention. For example, a transgenic plant expressing the late-flowering gene, *CONSTANS* (*CO*), flowers earlier than the corresponding wild type plant, but does not contain an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product (Putterill et al., *Cell* 80:847-857 (1995)). Thus, the early-flowering transgenic plant described by Putterill et al. is not a non-naturally occurring seed plant as defined herein.

Early reproductive development also has been observed in a transgenic tobacco plant expressing an exogenous rice MADS domain gene. Although the product of the rice MADS domain gene promotes early reproductive development, it does not specify the identity of floral meristem and, thus, cannot convert shoot meristem to floral meristem in an angiosperm (Chung et al., *Plant Mol. Biol.* 26:657-665 (1994)). Therefore, an early-flowering transgenic plant containing this rice MADS domain gene, like an early-flowering transgenic plant containing *CONSTANS*, is distinguishable from an early-flowering non-naturally occurring seed plant of the invention.

Mutations in a class of genes known as "early-flowering genes" also produce plants characterized by early reproductive development. Such early-flowering genes include, for example, *EARLY FLOWERING 1-3* (*ELF1*, *ELF2*, *ELF3*); *EMBRYONIC FLOWER 1,2* (*EMF1*, *EMF2*); *LONG HYPOCOTYL 1,2* (*HY1*, *HY2*); *PHYTOCHROME B* (*PHYB*), *SPINDLY* (*SPY*) and *TERMINAL FLOWER* (*TFL*) (Weigel, *supra*, 1995). The wild type product of an early-flowering gene retards reproductive development and is distinguishable from a floral meristem identity gene product in that an early-flowering gene product does not promote conversion of shoot meristem to floral meristem in an angiosperm. A plant that flowers early due to the loss of an early-flowering gene product function is distinct from a non-naturally occurring seed plant of the invention characterized by early reproductive development since such a plant does not contain an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product.

An *Arabidopsis* plant having a mutation in the *TERMINAL FLOWER* (*TFL*) gene is characterized by early reproductive development and by the conversion of shoots to flowers (Alvarez et al., *Plant J.* 2:103-116 (1992), which is incorporated herein by reference). However, *TFL* is not a floral meristem identity gene product, as defined herein. Specifically, it is the loss of *TFL* that promotes conversion of shoot meristem to floral meristem. Since the function of *TFL* is to antagonize formation of floral meristem, a *tfl* mutant, which lacks functional *TFL*, converts shoot meristem to floral meristem prematurely. Although *TFL* is

not a floral meristem identity gene product and does not itself convert shoot meristem to floral meristem, the loss of TFL can result in a plant with an ectopically expressed floral meristem identity gene product. However, such a *tfl* mutant, in which a mutation in an endogenous *TERMINAL FLOWER* gene results in conversion of shoot meristem to floral meristem, is excluded explicitly from the present invention.

In various embodiments, the present invention provides a non-naturally occurring seed plant containing a first ectopically expressible nucleic acid molecule encoding a first floral meristem identity gene product, provided that the first nucleic acid molecule is not ectopically expressed due to a mutation in an endogenous *TERMINAL FLOWER* gene. If desired, a non-naturally occurring seed plant of the invention can contain a second ectopically expressible nucleic acid molecule encoding SEP1, SE2, SEP3, AGL20, AGL22, AGL24, or AGL27, provided that the first or second nucleic acid molecule is not ectopically expressed due to a mutation in an endogenous *TERMINAL FLOWER* gene.

An ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product can be expressed, as desired, either constitutively or inducibly. Such an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product can be an endogenous floral meristem identity gene that has, for example, a mutation in a gene regulatory element. An ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product also can be an endogenous nucleic acid molecule encoding a floral meristem identity gene product that is linked to an exogenous, heterologous gene regulatory element that confers ectopic expression. In addition, an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product can be an exogenous nucleic acid molecule that encodes a floral meristem identity gene product under control of a heterologous gene regulatory element.

A non-naturally occurring seed plant of the invention can contain an endogenous floral meristem identity gene having a modified gene regulatory element. The term "modified gene regulatory element," as used herein in reference to the regulatory element of a floral meristem identity gene, means a regulatory element having a mutation that results in ectopic expression of the linked endogenous floral meristem identity gene. Such a gene regulatory element can be, for example, a promoter or enhancer element and can be positioned 5' or 3' to the coding sequence or within an intronic sequence of the floral meristem identity gene. A modified gene regulatory element can have, for example, a nucleotide insertion, deletion or substitution that is produced, for example, by chemical mutagenesis using a mutagen such as ethylmethane sulfonate or by insertional mutagenesis

using a transposable element. A modified gene regulatory element can be a functionally inactivated binding site for TFL or a functionally inactivated binding site for a gene product regulated by TFL, such that modification of the gene regulatory element results in ectopic expression of the linked floral meristem identity gene product, for example, in the shoot meristem of an angiosperm.

The present invention also provides a transgenic seed plant containing a first exogenous gene promoter that regulates a first ectopically expressible nucleic acid molecule encoding a first floral meristem identity gene product and a second exogenous gene promoter that regulates a second ectopically expressible nucleic acid molecule encoding a second floral meristem identity gene product.

The present invention further provides a transgenic seed plant containing a first exogenous ectopically expressible nucleic acid molecule encoding a first floral meristem identity gene product and a second exogenous gene promoter that regulates a second ectopically expressible nucleic acid molecule encoding a second floral meristem identity gene product, provided that the first nucleic acid molecule is not ectopically expressed due to a mutation in an endogenous *TERMINAL FLOWER* gene.

The invention also provides, therefore, a plant characterized by modulated (delayed or early) reproductive development, the plant containing a sense or antisense nucleic acid molecule encoding AP1, or a fragment thereof; a sense or antisense nucleic acid molecule encoding CAL, or a fragment thereof; and a sense or antisense nucleic acid molecule encoding LFY, or a fragment thereof, such that expression of AP1 and LFY gene products, including expression of endogenous AP1 and LFY gene products, is suppressed in the transgenic seed plant. Similarly, a sense or antisense nucleic acid molecule encoding SEP1, or a fragment thereof, a sense or antisense nucleic acid molecule encoding SEP2, or a fragment thereof, a sense or antisense nucleic acid molecule encoding SEP3, or a fragment thereof, a sense or antisense nucleic acid molecule encoding AGL20, or a fragment thereof, a sense or antisense nucleic acid molecule encoding AGL22, or a fragment thereof, a sense or antisense nucleic acid molecule encoding AGL24, or a fragment thereof, a sense or antisense nucleic acid molecule encoding AGL27, or a fragment thereof can also be used singly, in combination with each other or in combination with any of the AP1, CAL or LFY constructs discussed above.

In addition, the invention provides a transgenic seed plant containing a first exogenous ectopically expressible nucleic acid molecule encoding a first floral meristem identity gene product, provided that the first second nucleic acid molecule is not ectopically

expressed due to a mutation in an endogenous *TERMINAL FLOWER* gene, and further containing a second exogenous ectopically expressible nucleic acid molecule encoding a second floral meristem identity gene product, where the first floral meristem identity gene product is different from the second floral meristem identity gene product.

As disclosed herein, ectopic expression of two different floral meristem identity gene products can be particularly useful. For example, a fraction of the progeny of a cross between a transgenic *Arabidopsis* line constitutively expressing *AP1* under control of the cauliflower mosaic virus 35S promoter and a transgenic *Arabidopsis* line constitutively expressing *LFY* under control of the cauliflower mosaic virus 35S promoter are characterized by enhanced early reproductive development as compared to the early reproductive development of 35S-*AP1* transgenic lines or 35S-*LFY* transgenic lines. These results indicate that ectopic expression of the combination of *AP1* and *LFY* in a seed plant can result in enhanced early reproductive development as compared to the early reproductive development obtained by ectopic expression of *AP1* or *LFY* alone. Similarly, the ectopic expression of the combination of at least one of *AP1* and *CALIFLOWER* with at least one of *SEP1*, *SEP2*, *SEP3*, *AGL20*, *AGL22*, *AGL24* or *AGL27* results in early reproductive development. Thus, by using a combination of two different floral meristem identity gene products, plant breeding, for example, can be accelerated further as compared to the use of a single floral meristem identity gene product.

A useful combination of first and second floral meristem identity gene products can be, for example, *AP1* and *SEP3*, *CAL* and *SEP3*, *AP1* and *AGL24* or *CAL* and *AGL24*. Where a transgenic seed plant of the invention contains first and second exogenous nucleic acid molecules encoding different floral meristem identity gene products, it will be recognized that the order of introducing the first and second nucleic acid molecules into the seed plant is not important for purposes of the present invention. Thus, a transgenic seed plant of the invention having, for example, *AP1* as a first floral meristem identity gene product and *SEP3* as a second floral meristem identity gene product is equivalent to a transgenic seed plant having *SEP3* as a first floral meristem identity gene product and *AP1* as a second floral meristem identity gene product.

IV. Plant Transformation

As used herein, the term "introducing," when used in reference to a nucleic acid molecule and a seed plant such as an angiosperm or a gymnosperm, means transferring

an exogenous nucleic acid molecule into the seed plant. For example, an exogenous nucleic acid molecule encoding a floral meristem identity gene product can be introduced into a seed plant by a variety of methods including *Agrobacterium*-mediated transformation or direct gene transfer methods such as electroporation or microprojectile-mediated transformation.

Transformation methods based upon the soil bacterium *Agrobacterium tumefaciens*, known as "agro-infection," are useful for introducing a nucleic acid molecule into a broad range of angiosperms and gymnosperms. The wild type form of *Agrobacterium* contains a Ti (tumor-inducing) plasmid that directs production of tumorigenic crown gall growth on host plants. Transfer of the tumor-inducing T-DNA region of the Ti plasmid to a plant genome requires the Ti plasmid-encoded virulence genes as well as T-DNA borders, which are a set of direct DNA repeats that delineate the region to be transferred. *Agrobacterium*-based vector is a modified form of a Ti plasmid, in which the tumor inducing functions are replaced by nucleic acid sequence of interest to be introduced into the plant host.

Current protocols for *Agrobacterium*-mediated transformation employ cointegrate vectors or, preferably, binary vector systems in which the components of the Ti plasmid are divided between a helper vector, which resides permanently in the *Agrobacterium* host and carries the virulence genes, and a shuttle vector, which contains the gene of interest bounded by T-DNA sequences. A variety of binary vectors are well known in the art and are commercially available from, for example, Clontech (Palo Alto, California). Methods of coculturing *Agrobacterium* with cultured plant cells or wounded tissue such as leaf tissue, root explants, hypocotyledons, stem pieces or tubers, for example, also are well known in the art (Glick and Thompson (eds.), Methods in Plant Molecular Biology and Biotechnology, Boca Raton, FL: CRC Press (1993), which is incorporated herein by reference). Wounded cells within the plant tissue that have been infected by *Agrobacterium* can develop organs *de novo* when cultured under the appropriate conditions; the resulting transgenic shoots eventually give rise to transgenic plants containing the exogenous nucleic acid molecule of interest, as described in Example I.

Agrobacterium-mediated transformation has been used to produce a variety of transgenic seed plants (see, for example, Wang et al. (eds), Transformation of Plants and Soil Microorganisms, Cambridge, UK: University Press (1995), which is incorporated herein by reference). For example, *Agrobacterium*-mediated transformation can be used to produce transgenic crudiferous plants such as *Arabidopsis*, mustard, rapeseed and flax; transgenic leguminous plants such as alfalfa, pea, soybean, trefoil and white clover; and transgenic

solanaceous plants such as eggplant, petunia, potato, tobacco and tomato. In addition, *Agrobacterium*-mediated transformation can be used to introduce exogenous nucleic acids into apple, aspen, belladonna, black currant, carrot, celery, cotton, cucumber, grape, horseradish, lettuce, morning glory, muskmelon, neem, poplar, strawberry, sugar beet, sunflower, walnut and asparagus plants (see, for example, Glick and Thompson, *supra*, 1993).

Microprojectile-mediated transformation also is a well known method of introducing an exogenous nucleic acid molecule into a variety of seed plant species. This method, first described by Klein et al., *Nature* 327:70-73 (1987), which is incorporated herein by reference, relies on microprojectiles such as gold or tungsten that are coated with the desired nucleic acid molecule by precipitation with calcium chloride, spermidine or PEG. The microprojectile particles are accelerated at high speed into seed plant tissue using a device such as the BiolisticTM PD-1000 (Biorad, Hercules, California).

Microprojectile-mediated delivery or "particle bombardment" is especially useful to transform seed plants that are difficult to transform or regenerate using other methods. Microprojectile-mediated transformation has been used, for example, to generate a variety of transgenic seed plant species, including cotton, tobacco, corn, hybrid poplar and papaya (see, for example, Glick and Thompson, *supra*, 1993). The transformation of important cereal crops such as wheat, oat, barley, sorghum and rice also has been achieved using microprojectile-mediated delivery (Duan et al., *Nature Biotech.* 14:494-498 (1996); Shimamoto, *Curr. Opin. Biotech.* 5:158-162 (1994), each of which is incorporated herein by reference). A rapid transformation regeneration system for the production of transgenic plants, such as transgenic wheat, in two to three months also can be useful in producing a transgenic seed plant of the invention (European Patent No. EP 0 709 462 A2, Application number 95870117.9, filed 25 October 1995, which is incorporated herein by reference).

Thus, a variety of methods for introducing a nucleic acid molecule into a seed plant are well known in the art. Important crop species such as rice, for example, have been transformed using microprojectile delivery, *Agrobacterium*-mediated transformation or protoplast transformation (Hiei et al., *The Plant J.* 6(2):271-282 (1994); Shimamoto, *Science* 270:1772-1773 (1995), each of which is incorporated herein by reference). Fertile transgenic maize has been obtained, for example, by microparticle bombardment (see Wang et al., *supra*, 1995). As discussed above, barley, wheat, oat and other small-grain cereal crops also have been transformed, for example, using microparticle bombardment (see Wang et al., *supra*, 1995).

Methods of transforming forest trees including both angiosperms and gymnosperms also are well known in the art. Transgenic angiosperms such as members of the genus *Populus*, which includes aspens and poplars, have been generated using *Agrobacterium*-mediated transformation, for example. In addition, transgenic *Populus* and sweetgum, which are of interest for biomass production for fuel, also have been produced. Transgenic gymnosperms, including conifers such as white spruce and larch, also have been obtained, for example, using microprojectile bombardment (Wang et al., *supra*, 1995). The skilled artisan will recognize that *Agrobacterium*-mediated or microprojectile-mediated transformation, as disclosed herein, or other methods known in the art can be used to introduce a nucleic acid molecule encoding a floral meristem identity gene product into a seed plant according to the methods of the invention.

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pennisetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, and *Zea*.

V. Converting Shoot Meristem to Floral Meristem

The term "converting shoot meristem to floral meristem," as used herein, means promoting the formation of flower progenitor tissue where shoot progenitor tissue otherwise would be formed in the angiosperm. As a result of the conversion of shoot meristem to floral meristem, flowers form in an angiosperm where shoots normally would form. The conversion of shoot meristem to floral meristem can be identified using well known methods, such as scanning electron microscopy, light microscopy or visual inspection (see, for example, Mandel and Yanofsky, *Plant Cell* 7:1763-1771 (1995), which is incorporated herein by reference or Weigel and Nilsson, *supra*, 1995).

Provided herein are methods of converting shoot meristem to floral meristem in an angiosperm by introducing a first ectopically expressible nucleic acid molecule encoding a first floral meristem identity gene product and a second ectopically expressible

nucleic acid molecule encoding a second floral meristem identity gene product into the angiosperm, where the first floral meristem identity gene product is different from the second floral meristem identity gene product. As discussed above, first and second floral meristem identity gene products useful in converting shoot meristem to floral meristem in an angiosperm can be, for example, AP1 and LFY, CAL and LFY, or AP1 and CAL. In other preferred embodiments, the ectopic expression of the combination of at least one of AP1 and CALIFLOWER with at least one of SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27 results in conversion of shoot meristem to floral meristem.

VI. Methods of Modulating Reproductive Development

As discussed above, the present invention provides methods of promoting modulated timing of reproductive development in a seed plant by ectopically expressing a first nucleic acid molecule encoding a first floral meristem identity gene product in the seed plant, provided that the first nucleic acid molecule is not ectopically expressed due to a mutation in an endogenous *TERMINAL FLOWER* gene. For example, the invention provides a method of promoting modulated timing of reproductive development in a seed plant by introducing an ectopically expressible nucleic acid molecule encoding a floral meristem identity gene product into the seed plant, thus producing a transgenic seed plant. A floral meristem identity gene product such as SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, AGL27, AP1, CAL or LFY, or a chimeric protein containing, in part, a floral meristem identity gene product, as disclosed below, is useful in methods of promoting early reproductive development.

The term "promoting early reproductive development," as used herein in reference to a seed plant, means promoting the formation of a reproductive structure earlier than the time when a reproductive structure would form on a corresponding seed plant that is grown under the same conditions and that does not ectopically express a floral meristem identity gene product. As discussed above, the time when reproductive structures form on a particular seed plant that does not ectopically express a floral meristem identity gene product is relatively fixed and depends, in part, on genetic factors as well as environmental conditions, such as day length and temperature. Thus, given a defined set of environmental conditions, a naturally occurring angiosperm, for example, will flower at a relatively fixed time. Similarly, given a defined set of environmental conditions, a naturally occurring coniferous gymnosperm, for example, will produce cones at a relatively fixed time.

Methods for ectopically expressing polynucleotides in plants are well known in the art. For example, the expression of polynucleotides of the invention can be modulated by mutation, or introduction of at least one copy of the polynucleotides into a plant.

One of skill will recognize that a number of methods can be used to modulate gene product activity or gene expression. Gene product activity can be modulated in the plant cell at the gene, transcriptional, posttranscriptional, translational, or posttranslational, levels. Techniques for modulating gene product activity at each of these levels are generally well known to one of skill and are discussed briefly below. "Activity" encompasses both mechanistic activities (e.g., enzymatic, ability to induce transcription of genes under the gene products control, etc.) and phenotypic activities such as altering the time of reproductive development.

Methods for introducing genetic mutations into plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used.

Alternatively, homologous recombination can be used to induce targeted gene disruptions by specifically deleting or altering the target gene *in vivo* (see, generally, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al.*, *Genes Dev.* 10:2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al.*, *Experientia* 50:277-284 (1994), Swoboda *et al.*, *EMBO J.* 13:484-489 (1994); Offringa *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al.* *Nature* 389:802-803 (1997)).

In applying homologous recombination technology to the genes of the invention, mutations in selected portions of a gene sequences (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed here are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al.* *Proc. Natl. Acad. Sci. USA* 91:4303-4307 (1994); and Vaulont *et al.* *Transgenic Res.* 4:247-255 (1995) are conveniently used to increase the efficiency of selecting for altered gene expression in transgenic plants. The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in suppression of gene product activity.

Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used. The RNA/DNA sequence is designed to align with the sequence of the target gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an
5 extrachromosomal T-DNA plasmid results in efficient and specific gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is described in Cole-Strauss *et al. Science* 273:1386-1389 (1996) and Yoon *et al. Proc. Natl. Acad. Sci. USA* 93:2071-2076 (1996).

Gene expression can be inactivated using recombinant DNA techniques by
10 transforming plant cells with constructs comprising transposons or T-DNA sequences. Mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of mRNA, *e.g.*, by Northern blots. Mutants can also be selected by assaying for altered timing of the development of reproductive structures.

15 The isolated nucleic acid sequences prepared as described herein, can also be used in a number of techniques to control endogenous gene expression at various levels. Subsequences from the sequences disclosed here can be used to control, transcription, RNA accumulation, translation, and the like.

A number of methods can be used to inhibit gene expression in plants. For
20 instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense suppression can act at all levels of gene regulation including suppression of RNA
25 translation (*see, Bourque Plant Sci. (Limerick)* 105:125-149 (1995); Pantopoulos In Progress in Nucleic Acid Research and Molecular Biology, Vol. 48. Cohn, W. E. and K. Moldave (Ed.). Academic Press, Inc.: San Diego, California, USA; London, England, UK. p. 181-238; Heiser *et al. Plant Sci. (Shannon)* 127:61-69 (1997)) and by preventing the accumulation of mRNA which encodes the protein of interest, (*see, Baulcombe Plant Mol. Bio.* 32:79-88
30 (1996); Prins and Goldbach *Arch. Virol.* 141:2259-2276 (1996); Metzlaiff *et al. Cell* 88:845-854 (1997), Sheehy *et al., Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340).

The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous gene or genes to be repressed. The sequence,

however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 7000 nucleotides is especially preferred.

A number of gene regions can be targeted to suppress gene expression. The targets can include, for instance, the coding regions, introns, sequences from exon/intron junctions, 5' or 3' untranslated regions, and the like. In some embodiments, the constructs can be designed to eliminate the ability of regulatory proteins to bind to gene sequences that are required for its cell- and/or tissue-specific expression. Such transcriptional regulatory sequences can be located either 5'-, 3'-, or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription. These sequences can be identified using standard deletion analysis, well known to those of skill in the art. Once the sequences are identified, an antisense construct targeting these sequences is introduced into plants to control gene transcription in particular tissue, for instance, in developing ovules and/or seed. In one embodiment, transgenic plants are selected for activity that is reduced but not eliminated.

Oligonucleotide-based triple-helix formation can be used to disrupt gene expression. Triplex DNA can inhibit DNA transcription and replication, generate site-specific mutations, cleave DNA, and induce homologous recombination (*see, e.g.,* Havre and Glazer *J. Virology* 67:7324-7331 (1993); Scanlon *et al. FASEB J.* 9:1288-1296 (1995); Giovannangeli *et al. Biochemistry* 35:10539-10548 (1996); Chan and Glazer *J. Mol. Medicine (Berlin)* 75:267-282 (1997)). Triple helix DNAs can be used to target the same sequences identified for antisense regulation.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered,

and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. Thus, ribozymes can be used to target the same sequences identified for antisense regulation.

5 A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco
10 mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Zhao and Pick *Nature* 365:448-451 (1993); Eastham and Ahlering *J. Urology* 156:1186-1188 (1996); Sokol and Murray *Transgenic Res.* 5:363-371 (1996); Sun *et al. Mol. Biotechnology* 7:241-251 (1997); and Haseloff *et al. Nature*, 334:585-591 (1988).

15 Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes (*see*, Assaad *et al. Plant Mol. Bio.* 22:1067-1085 (1993); Flavell *Proc. Natl. Acad. Sci. USA* 91:3490-3496 (1994); Stam *et al.*
20 *Annals Bot.* 79:3-12 (1997); Napoli *et al., The Plant Cell* 2:279-289 (1990); and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184).

 The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence
25 generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other
30 proteins within a similar family of genes exhibiting homology or substantial homology.

 For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants that are overexpressers. A higher identity in a shorter than full length sequence compensates

for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used. In addition, the same gene regions noted for antisense regulation can be targeted using
5 cosuppression technologies.

In a preferred embodiment, expression of a nucleic acid of interest can be suppressed by the simultaneous expression of both sense and antisense constructs (Waterhouse *et al.*, *Proc. Natl. Acad. Sci. USA* 95:13959-13964 (1998). *See also* Tabara *et al. Science* 282:430-431 (1998).

10 Alternatively, gene product activity may be modulated by eliminating the proteins that are required for cell-specific gene expression. Thus, expression of regulatory proteins and/or the sequences that control gene expression can be modulated using the methods described here.

Another method is use of engineered tRNA suppression of mRNA translation.
15 This method involves the use of suppressor tRNAs to transactivate target genes containing premature stop codons (*see*, Betzner *et al. Plant J.* 11:587-595 (1997); and Choisine *et al. Plant J.* 11:597-604 (1997). A plant line containing a constitutively expressed gene that contains an amber stop codon is first created. Multiple lines of plants, each containing tRNA suppressor gene constructs under the direction of cell-type specific promoters are also
20 generated. The tRNA gene construct is then crossed into the desired gene product line to activate activity in a targeted manner. These tRNA suppressor lines could also be used to target the expression of any type of gene to the same cell or tissue types.

Proteins may form homogeneous or heterologous complexes *in vivo*. Thus, production of dominant-negative forms of polypeptides that are defective in their abilities to
25 bind to other proteins in the complex is a convenient means to inhibit endogenous gene product activity. This approach involves transformation of plants with constructs encoding mutant polypeptides that form defective complexes and thereby prevent the complex from forming properly. The mutant polypeptide may vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like.
30 These modifications can be used in a number of combinations to produce the final modified protein chain. Use of dominant negative mutants to inactivate target genes is described in Mizukami *et al. Plant Cell* 8:831-845 (1996).

Another strategy to affect the ability of a protein to interact with itself or with other proteins involves the use of antibodies specific to the protein. In this method cell-

specific expression of specific Abs is used inactivate functional domains through antibody:antigen recognition (*see, Hupp et al. Cell* 83:237-245 (1995)).

After plants with reduced activity are identified, a recombinant construct capable of expressing low levels of the gene product can be introduced using the methods discussed below. In this fashion, the level of activity can be regulated to produce preferred plant phenotypes. For example, a relatively weak promoter such as the ubiquitin promoter (see, e.g., Garbarino *et al. Plant Physiol.* 109(4):1371-8 (1995); Christensen *et al Transgenic Res.* 5(3):213-8 (1996); and Holtorf *et al. Plant. Mol. Biol.* 29(4):637-46 (1995)) is useful to produce plants with reduced levels of activity or expression. Such plants are useful for producing, for instance, plants with altered time of developing reproductive structures.

As disclosed herein, ectopic expression of a nucleic acid molecule encoding a floral meristem identity gene product in an angiosperm converts shoot meristem to floral meristem in the angiosperm. Furthermore, ectopic expression of a nucleic acid molecule encoding a floral meristem identity gene product such as AP1, CAL or LFY in an angiosperm prior to the time when endogenous floral meristem identity gene products are expressed in the angiosperm can convert shoot meristem to floral meristem precociously, resulting in early reproductive development in the angiosperm, as indicated by early flowering. In the same manner, ectopic expression of a nucleic acid molecule encoding AP1, CAL, or LFY, for example, in a gymnosperm prior to the time when endogenous floral meristem identity gene products are expressed in the gymnosperm results in early reproductive development in the gymnosperm.

For a given seed plant species and particular set of growth conditions, constitutive expression of a floral meristem identity gene product results in a relatively invariant time of early reproductive development, which is the earliest time when all factors necessary for reproductive development are active. For example, constitutive expression of *AP1* in transgenic *Arabidopsis* plants grown under "long-day" light conditions results in early reproductive development at day 10 as compared to the normal time of reproductive development, which is day 18 in non-transgenic *Arabidopsis* plants grown under the same conditions. Thus, under these conditions, day 10 is the relatively invariant time of early reproductive development for *Arabidopsis* transgenics that constitutively express a floral meristem identity gene product. Similarly, transgenic plants constitutively expressing *SEP3* result in plants that develop earlier reproductive structures than wild type plants.

However, in addition to methods of constitutively expressing a floral meristem identity gene product, the present invention provides methods of selecting the time of early

reproductive development. As disclosed herein, floral meristem gene product expression or activity can be regulated in response to an inducing agent or cognate ligand, for example, such that the time of reproductive development can be selected. For example, in *Arabidopsis* transgenics grown under the conditions described above, the time of early reproductive development need not necessarily be the relatively invariant day 10 at which early reproductive development occurs as a consequence of constitutive floral meristem identity gene product expression. If floral meristem identity gene product expression is rendered dependent upon the presence of an inducing agent, early reproductive development can be selected to occur, for example, on day 14, by contacting the seed plant with an inducing agent on or slightly before day 14.

Thus, the present invention provides recombinant nucleic acid molecules, transgenic seed plant containing such recombinant nucleic acid molecules and methods for selecting the time of early reproductive development. These methods allow a farmer or horticulturist, for example, to determine the time of early reproductive development. The methods of the invention can be useful, for example, in allowing a grower to respond to an approaching storm or impending snap-freeze by selecting the time of early reproductive development such that the crop can be harvested before being harmed by the adverse weather conditions. The methods of the invention for selecting the time of early reproductive development also can be useful to spread out the time period over which transgenic seed plants are ready to be harvested. For example, the methods of the invention can be used to increase floral meristem identity gene product expression in different crop fields at different times, resulting in a staggered time of harvest for the different fields.

Thus, the present invention provides a recombinant nucleic acid molecule containing an inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product. The floral meristem identity gene product encoded within a recombinant nucleic acid molecule of the invention can be, for example, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, AGL27, AP1 or CAL. In addition, the floral meristem identity gene product encoded within a recombinant nucleic acid molecule of the invention can be LFY. As disclosed herein, a recombinant nucleic acid molecule of the invention can contain an inducible regulatory element such as a copper inducible element, tetracycline inducible element, ecdysone inducible element or heat shock inducible element.

VII. Inducible Regulatory Elements

The invention also provides a transgenic seed plant containing a recombinant nucleic acid molecule comprising an inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product. Such a transgenic seed plant can be an angiosperm or gymnosperm and can contain, for example, a recombinant nucleic acid molecule comprising an inducible regulatory element operably linked to a nucleic acid molecule encoding AP1 or CAL. Similarly, the ectopic expression of the combination of at least one of AP1 and CALIFLOWER with at least one of SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27 can be used to produce seed with various desirable phenotypes. A transgenic seed plant of the invention can contain, for example, a recombinant nucleic acid molecule comprising a copper inducible element, tetracycline inducible element, ecdysone inducible element or heat shock inducible element operably linked to a nucleic acid molecule encoding AP1, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27. In addition, a transgenic seed plant of the invention can contain a recombinant nucleic acid molecule comprising a copper inducible element tetracycline inducible element, ecdysone inducible element or heat shock inducible element operably linked to a nucleic acid molecule encoding CAL. A transgenic seed plant of the invention also can contain a recombinant nucleic acid molecule comprising a copper inducible element, tetracycline inducible element, ecdysone inducible element or heat shock inducible element operably linked to a nucleic acid molecule encoding LFY.

A particularly useful inducible regulatory element can be, for example, a copper-inducible promoter (Mett et al., *Proc. Natl. Acad. Sci. USA* 90:4567-4571 (1993), which is incorporated herein by reference); tetracycline-inducible regulatory element (Gatz et al., *Plant J.* 2:397-404 (1992); Röder et al., *Mol. Gen. Genet.* 243:32-38 (1994), each of which is incorporated herein by reference); ecdysone inducible element (Christopherson et al., *Proc. Natl. Acad. Sci. USA* 89:6314-6318 (1992), which is incorporated herein by reference); or heat shock inducible element (Takahashi et al., *Plant Physiol.* 99:383-390 (1992), which is incorporated herein by reference). Another useful inducible regulatory element can be a lac operon element, which is used in combination with a constitutively expressed lac repressor to confer, for example, IPTG-inducible expression, as described by Wilde et al., (*EMBO J.* 11:1251-1259 (1992), which is incorporated herein by reference).

An inducible regulatory element useful in a method of the invention also can be, for example, a nitrate-inducible promoter derived from the spinach nitrite reductase gene

(Back et al., *Plant Mol. Biol.* 17:9 (1991), which is incorporated herein by reference) or a light-inducible promoter, such as that associated with the small subunit of RuBP carboxylase or the LHCP gene families (Feinbaum et al., *Mol. Gen. Genet.* 226:449 (1991); Lam and Chua, *Science* 248:471 (1990), each of which is incorporated herein by reference). An
5 inducible regulatory element useful in constructing a transgenic seed plant also can be a salicylic acid inducible element (Uknes et al., *Plant Cell* 5:159-169 (1993); Bi et al., *Plant J.* 8:235-245 (1995), each of which is incorporated herein by reference) or a plant hormone-inducible element (Yamaguchi-Shinozaki et al., *Plant Mol. Biol.* 15:905 (1990); Kares et al., *Plant Mol. Biol.* 15:225 (1990), each of which is incorporated herein by
10 reference). A human glucocorticoid response element also is an inducible regulatory element that can confer hormone-dependent gene expression in seed plants (Schena et al., *Proc. Natl. Acad. Sci. USA* 88:10421 (1991), which is incorporated herein by reference).

An inducible regulatory element that is particularly useful for increasing expression of a floral meristem identity gene product in a transgenic seed plant of the
15 invention is a copper inducible regulatory element (see, for example, Mett et al., *supra*, 1993). Thus, the invention provides a recombinant nucleic acid molecule comprising a copper inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product and a transgenic seed plant containing such a recombinant nucleic acid molecule. Copper, which is a natural part of the nutrient
20 environment of a seed plant, can be used to increase expression of a nucleic acid molecule encoding a floral meristem identity gene product operably linked to a copper inducible regulatory element. For example, an ACE1 binding site in conjunction with constitutively expressed yeast ACE1 protein confers copper inducible expression upon an operably linked nucleic acid molecule. The ACE1 protein, a metalloresponsive transcription factor, is
25 activated by copper or silver ions, resulting in increased expression of a nucleic acid molecule operably linked to an ACE1 element.

Such a copper inducible regulatory element can be an ACE1 binding site from the metallothionein gene promoter (SEQ ID NO: 21; Furst et al., *Cell* 55:705-717 (1988), which is incorporated herein by reference). For example, the ACE1 binding site can be
30 combined with the 90 base-pair domain A of the cauliflower mosaic virus 35S promoter and operably linked to a nucleic acid molecule encoding AP1, CAL or LFY to produce a recombinant nucleic acid molecule of the invention. In a transgenic seed plant constitutively expressing ACE1 under control of such a modified CaMV 35S promoter, for example, copper

inducible expression is conferred upon an operably linked nucleic acid molecule encoding a floral meristem identity gene product.

The expression of a nucleic acid encoding a floral meristem identity gene product operably linked to a copper inducible regulatory element, such as

5 5'-AGCTTAGCGATGCGTCTTTTCCGCTGAACCGTTCCAGCAAAAAGACTAG-3'
(SEQ ID NO: 21), can be increased in a transgenic seed plant grown under copper ion-depleted conditions, for example, and contacted with 50 μ M copper sulfate in a nutrient solution or with 0.5 μ M copper sulfate applied by foliar spraying of the transgenic seed plant (see, for example, Mett et al., *supra*, 1993). A single application of 0.5 μ M copper sulfate
10 can be sufficient to sustain increased floral meristem identity gene product expression over a period of several days. If desired, a transgenic seed plant of the invention also can be contacted with multiple applications of an inducing agent such as copper sulfate.

An inducible regulatory element also can confer tetracycline-dependent floral meristem identity gene expression in a transgenic seed plant of the invention. Thus, the
15 present invention provides a recombinant nucleic acid molecule comprising a tetracycline inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product as well as a transgenic seed plant into which such a recombinant nucleic acid molecule has been introduced. A tetracycline inducible regulatory element is particularly useful for conferring tightly regulated gene expression as indicated by
20 the observation that a phenotype that results from even low amounts of a gene product expression is suppressed from such an inducible system in the absence of inducing agent (see, for example, Röder et al., *supra*, 1994).

A transgenic seed plant constitutively expressing Tn10-encoded Tet repressor (TetR), for example, can be contacted with tetracycline to increase expression of a nucleic
25 acid molecule encoding a floral meristem identity gene product operably linked to the cauliflower mosaic virus promoter containing several *tet* operator sequences (5'-ACTCTATCAGTGATAGAGT-3'; SEQ ID NO: 22) positioned close to the TATA box (see, for example, Gatz, *Meth. Cell Biol.* 50:411-424 (1995), which is incorporated herein by reference; Gatz et al., *supra*, 1992). Such a tetracycline-inducible system can increase
30 expression of an operably linked nucleic acid molecule as much as 200 to 500-fold in a transgenic angiosperm or gymnosperm of the invention.

A high level of Tet repressor expression (about 1×10^6 molecules per cell) is critical for tight regulation. Thus, a seed plant preferably is transformed first with a plasmid encoding the Tet repressor, and screened for high level expression. For example, plasmid

pBinTet (Gatz, *supra*, 1995) contains the Tet repressor coding region, which is expressed under control of the CaMV 35S promoter, and the neomycin phosphotransferase gene for selection of transformants. To screen transformants for a high level of Tet repressor expression, a plasmid containing a reporter gene under control of a promoter with *tet* operators, such as pTX-Gus-int (Gatz, *supra*, 1995), can be transiently introduced into a seed plant cell and assayed for activity in the presence and absence of tetracycline. High β -glucouronidase (GUS) expression that is dependent on the presence of tetracycline is indicative of high Tet repressor expression.

A particularly useful tetracycline inducible regulatory element is present in plasmid pBIN-HygTX, which has a CaMV 35S promoter, into which three tet operator sites have been inserted, and an octopine synthase polyadenylation site (Gatz, *supra*, 1995). A multiple cloning site between the promoter and polyadenylation signal in pBIN-HygTX allows for convenient insertion of a nucleic acid molecule encoding the desired floral meristem identity gene product, and the hygromycin phosphotransferase gene allows for selection of transformants containing the construct. In a preferred embodiment of the invention, previously selected Tet repressor positive cells are transformed with a plasmid such as pBIN-HygTX, into which a nucleic acid molecule encoding a floral meristem identity gene product has been inserted.

To increase floral meristem identity gene product expression using a tetracycline-inducible regulatory element, a transgenic seed plant of the invention can be contacted with tetracycline or, preferably, with chlor-tetracycline (SIGMA), which is a more efficient inducer than tetracycline. In addition, a useful inducing agent can be a tetracycline analog that binds the Tet repressor to function as an inducer but that does not act as an antibiotic (Gatz, *supra*, 1995). A transgenic seed plant of the invention can be contacted, for example, by watering with about 1 mg/liter chlor-tetracycline or tetracycline. Similarly, a plant grown in hydroponic culture can be contacted with a solution containing about 1 mg/liter chlor-tetracycline or tetracycline (Gatz, *supra*, 1995). If desired, a transgenic angiosperm or gymnosperm can be contacted repeatedly with chlor-tetracycline or tetracycline every other day for about 10 days (Röder et al., *supra*, 1994). Floral meristem identity gene product expression is increased efficiently at a tetracycline concentration that does not inhibit the growth of bacteria, indicating that the use of tetracycline as an inducing agent will not present environmental concerns.

An ecdysone inducible regulatory element also can be useful in practicing the methods of the invention. For example, an ecdysone inducible regulatory element can contain four copies of an ecdysone response element having the sequence 5'-GATCCGACAAGGGTTCAATGCACTTGTCA-3' (EcRE; SEQ ID NO: 23) as described in Christopherson et al., *supra*, 1992. In a transgenic seed plant into which a nucleic acid encoding an ecdysone receptor has been introduced, an ecdysone inducible regulatory element can confer ecdysone-dependent expression on a nucleic acid molecule encoding a floral meristem identity gene product. An appropriate inducing agent for increasing expression of a nucleic acid molecule operably linked to an ecdysone inducible regulatory element can be, for example, ∇ -ecdysone, 20-hydroxyecdysone, polypodine B, ponasterone A, muristerone A or RH-5992, which is an ecdysone agonist that mimics 20-hydroxyecdysone (see, for example, Kreutzweiser et al., *Ecotoxicol. Environ. Safety* 28:14-24 (1994), which is incorporated herein by reference and Christopherson et al., *supra*, 1992). Methods for determining an appropriate inducing agent for use with an ecdysone inducible regulatory element are well known in the art. As disclosed herein, compound RH-5992 can be a particularly useful inducing agent for increasing floral meristem gene product expression in a transgenic seed plant containing an ecdysone inducible regulatory element.

An inducible regulatory element also can be derived from the promoter of a heat shock gene, such as *HSP81-1* (SEQ ID NO: 24; Takahashi, *supra*, 1992). Thus, the invention also provides a recombinant nucleic acid molecule comprising a heat shock inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product and a transgenic seed plant containing such a recombinant nucleic acid molecule. The *HSP81-1* promoter (SEQ ID NO: 24) confers low level expression upon an operably linked nucleic acid molecule in parts of roots under unstressed conditions and confers high level expression in most *Arabidopsis* tissues following heat shock (see, for example, Yabe et al., *Plant Cell Physiol.* 35:1207-1219 (1994), which is incorporated herein by reference). After growth of *Arabidopsis* at 23EC, a single heat shock treatment at 37EC for two hours is sufficient to induce expression of a nucleic acid molecule operably linked to the *HSP81-1* gene regulatory element (see Ueda et al., *Mol. Gen. Genet.* 250:533-539 (1996), which is incorporated herein by reference).

The use of a heat shock inducible regulatory element is particularly useful for a transgenic seed plant of the invention grown in an enclosed environment such as a green

house, where temperature can be readily manipulated. The use of a heat shock inducible regulatory element especially is applicable to a transplantable or potted transgenic seed plant of the invention, which can be moved conveniently from an environment having a low temperature to an environment having a high temperature. A transgenic angiosperm or gymnosperm of the invention containing a recombinant nucleic acid molecule comprising a *HSP81-1* heat shock regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product also can be induced, for example, by altering the ambient temperature, watering with heated water or submersing the transgenic seed plant in a sealed plastic bag into a heated water bath (see, for example, Ueda et al., *supra*, 1996).

A recombinant nucleic acid molecule of the invention comprising an inducible gene regulatory element can be expressed variably in different lines of transgenic seed plants. In some transgenic lines, for example, leaky expression of the introduced recombinant nucleic acid molecule can occur in the absence of the appropriate inducing agent due to phenomena such as position effects (see, for example, Ueda et al., *supra*, 1996). Thus, a transgenic seed plant containing a recombinant nucleic acid molecule comprising an inducible gene regulatory element operably linked to a nucleic acid encoding a floral meristem identity gene product can be screened, if desired, to obtain a particular transgenic seed plant in which expression of the operably linked nucleic acid molecule is desirably low in the absence of the appropriate inducing agent.

The present invention also provides a method of converting shoot meristem to floral meristem in an angiosperm by introducing into the angiosperm a recombinant nucleic acid molecule comprising an inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product to produce a transgenic angiosperm, and contacting the transgenic angiosperm with an inducing agent, thereby increasing expression of the floral meristem identity gene product and converting shoot meristem to floral meristem in the transgenic angiosperm. In such a method of the invention, the inducible regulatory element can be, for example, a copper inducible element, tetracycline inducible element, ecdysone inducible element or heat shock inducible element, and the floral meristem identity gene product can be, for example, AP1, CAL, LFY, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27.

In addition, the invention provides a method of promoting early reproductive development in a seed plant such as an angiosperm or gymnosperm by introducing into the seed plant a recombinant nucleic acid molecule comprising an inducible regulatory element operably linked to a nucleic acid molecule encoding a floral meristem identity gene product

to produce a transgenic seed plant, and contacting the transgenic seed plant with an inducing agent, thereby increasing expression of the floral meristem identity gene product and promoting early reproductive development in the transgenic seed plant. In a method of the invention for promoting early reproductive development in a seed plant, the inducible regulatory element can be, for example, a copper inducible element, tetracycline inducible element, ecdysone inducible element or heat shock inducible element, and the floral meristem identity gene product can be, for example, AP1, CAL, LFY, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27.

The term "inducing agent," as used herein, means a substance or condition that effects increased expression of a nucleic acid molecule operably linked to a particular inducible regulatory element as compared to the level of expression of the nucleic acid molecule in the absence of the inducing agent. An inducing agent can be, for example, a naturally occurring or synthetic chemical or biological molecule such as a simple or complex organic molecule, a peptide, a protein or an oligonucleotide that increases expression of a nucleic acid molecule operably linked to a particular inducible regulatory element. An example of such an inducing agent is a compound such as copper sulfate, tetracycline or an ecdysone. An inducing agent also can be a condition such as heat of a certain temperature or light of a certain wavelength. When used in reference to a particular inducible regulatory element, an "appropriate" inducing agent means an inducing agent that results in increased expression of a nucleic acid molecule operably linked to the particular inducible regulatory element.

An inducing agent of the invention can be used alone or in solution or can be used in conjunction with an acceptable carrier that can serve to stabilize the inducing agent or to promote absorption of the inducing agent by a seed plant. If desired, a transgenic seed plant of the invention can be contacted with an inducing agent in combination with an unrelated substance such as a plant nutrient, pesticide or insecticide.

One skilled in the art can readily determine the optimum concentration of an inducing agent needed to produce increased expression of a nucleic acid molecule operably linked to an inducible regulatory element in a transgenic seed plant of the invention. For conveniently determining the optimum concentration of inducing agent from a range of useful concentrations, one skilled in the art can operably link the particular inducible regulatory element to a nucleic acid molecule encoding a reporter gene product such as β -glucouronidase (GUS) and assay for reporter gene product activity in the presence of

various concentrations of inducing agent (see, for example, Jefferson et al., *EMBO J.* 6:3901-3907 (1987), which is incorporated herein by reference).

As used herein, the term "contacting," in reference to a transgenic seed plant of the invention, means exposing the transgenic seed plant to an inducing agent, or to a cognate ligand as disclosed below, such that the agent can induce expression of a nucleic acid molecule operably linked to the particular inducible regulatory element. A transgenic seed plant such as an angiosperm or gymnosperm, which contains a recombinant nucleic acid molecule of the invention, can be contacted with an inducing agent in a variety of manners. Expression of a floral meristem identity gene product can be increased conveniently, for example, by spraying a transgenic seed plant with an aqueous solution containing an appropriate inducing agent or by adding an appropriate inducing agent to the water supply of a transgenic seed plant grown using irrigation or to the water supply of a transgenic seed plant grown hydroponically. A transgenic seed plant containing a recombinant nucleic acid molecule of the invention also can be contacted by spraying the seed plant with an inducing agent in aerosol form. In addition, a transgenic seed plant can be contacted with an appropriate inducing agent by adding the agent to the soil or other solid nutrient media in which the seed plant is grown, whereby the inducing agent is absorbed into the seed plant. Other modes of contacting a transgenic seed plant with an inducing agent, such as injecting or immersing the seed plant in a solution containing an inducing agent, are well known in the art. For an inducing agent that is temperature or light, for example, contacting can be effected by altering the temperature or light to which the transgenic seed plant is exposed, or, if desired, by moving the transgenic seed plant from an environment of one temperature or light source to an environment having the appropriate inducing temperature or light source.

If desired, a transgenic seed plant of the invention can be contacted individually with an inducing agent. Furthermore, a group of transgenic seed plants that, for example, are located together in a garden plot, hot house or field, can be contacted *en masse* with an inducing agent, such that floral meristem identity gene product expression is increased coordinately in all transgenic seed plants of the group.

A transgenic seed plant of the invention can be contacted with an inducing agent using one of several means. For example, a transgenic seed plant can be contacted with an inducing agent by non-automated means such as with a hand held spraying apparatus. Such manual means can be useful when the methods of the invention are applied to particularly delicate or valuable seed plant varieties or when it is desirable, for example, to promote early reproductive development in a particular transgenic seed plant without

promoting early reproductive development in a neighboring transgenic seed plant.

Furthermore, a transgenic seed plant of the invention can be contacted with an inducing agent by mechanical means such as with a conventional yard "sprinkler" for a transgenic seed plant grown, for example, in a garden; a mechanical spraying system in a green house; traditional
5 farm machinery for spraying field crops; or "crop dusting" for conveniently contacting an entire field of transgenic seed plants with a particulate or gaseous inducing agent. The skilled practitioner, whether home gardener or commercial farmer, recognizes that these and other manual or mechanical means can be used to contact a transgenic seed plant with an inducing agent according to the methods of the invention.

10 Furthermore, it is recognized that a transgenic seed plant of the invention can be contacted with a single treatment of an inducing agent or, if desired, can be contacted with multiple applications of the inducing agent. In a preferred embodiment of the invention, a transgenic seed plant of the invention is contacted once with an inducing agent to effectively increase floral meristem identity gene product expression, thereby promoting early
5 reproductive development in the transgenic seed plant. Similarly, a transgenic angiosperm of the invention preferably is contacted once with an inducing agent to effectively increase floral meristem identity gene product expression and convert shoot meristem to floral meristem in the transgenic angiosperm.

A single application of an inducing agent is preferable when a transient
20 increase in floral meristem identity gene product expression from a recombinant nucleic acid molecule of the invention promotes irreversible early reproductive development in a seed plant. In many seed plant species, early reproductive development is irreversible. Transient expression of a floral meristem identity gene product from an introduced recombinant nucleic acid molecule, for example, results in sustained ectopic expression of endogenous floral
25 meristem identity gene products, resulting in irreversible early reproductive development. For example, ectopic expression of AP1 in a transgenic plant induces endogenous LFY gene expression, and ectopic expression of LFY induces endogenous AP1 gene expression (Mandel and Yanofsky, *Nature* 377:522-524 (1995), which is incorporated herein by reference; Weigel and Nilsson, *supra*, 1995). Genetic studies also indicate that CAL can act
30 directly or indirectly to increase expression of AP1 and LFY. Thus, ectopic expression of CAL from an exogenous nucleic acid molecule, for example, can induce endogenous AP1 and LFY expression (see Bowman et al., *supra*, 1993). Enhanced expression of endogenous AP1, LFY or CAL following a transient increase in expression of an introduced floral

meristem identity gene product induced by a single application of an inducing agent can make repeated applications of an inducing agent unnecessary.

In some seed plants, however, such as angiosperms characterized by the phenomenon of floral reversion, repeated applications of the inducing agent can be desirable.

5 In species such as impatiens, an initiated flower can revert into a shoot such that the center of the developing flower behaves as an indeterminate shoot (see, for example, Battey and Lyndon, *Ann. Bot.* 61:9-16 (1988), which is incorporated by reference herein). Thus, to prevent floral reversion in species such as impatiens, repeated applications of an inducing agent can be useful. Repeated applications of an inducing agent, as well as single
10 applications, are encompassed within the scope of the present invention.

VIII. Chimeric Polypeptides of the Invention

The invention further provides a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding a floral meristem identity gene product such as SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, AGL27, AP1, CAL or LFY
15 linked in frame to a nucleic acid molecule encoding a ligand binding domain. Expression of a chimeric protein of the invention in a seed plant is useful because the ligand binding domain renders the activity of a linked gene product dependent on the presence of cognate ligand. Specifically, in a chimeric protein of the invention, floral meristem gene product activity is increased in the presence of cognate ligand, as compared to activity in the absence
20 of cognate ligand.

A nucleic acid molecule encoding a chimeric protein of the invention comprises a nucleic acid molecule encoding a floral meristem identity gene product, such as a nucleic acid molecule having the nucleic acid sequence SEQ ID NO: 1, SEQ ID NO: 9, SEQ
25 ID NO: 15, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, which encode AP1, CAL, LFY, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 and AGL27, respectively, any of which is linked in frame to a nucleic acid molecule encoding a ligand binding domain. The expression of such a nucleic acid molecule results in the production of a chimeric protein containing a floral meristem
30 identity gene product fused to a ligand binding domain. Thus, the invention also provides a chimeric protein containing a floral meristem identity gene product fused to a ligand binding domain and an antibody that specifically binds such a chimeric protein.

The invention further provides a transgenic seed plant, such as angiosperm or gymnosperm, that contains a nucleic acid molecule encoding a chimeric protein of the invention. The invention provides, for example, a transgenic seed plant containing a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule
5 encoding AP1, CAL or LFY linked in frame to a nucleic acid molecule encoding a ligand binding domain. A particularly useful transgenic seed plant contains a nucleic acid molecule encoding AP1 linked in frame to a nucleic acid molecule encoding an ecdysone receptor ligand binding domain or a glucocorticoid receptor ligand binding domain. The invention also provides a transgenic seed plant containing a nucleic acid molecule encoding a chimeric
10 protein, which comprises a nucleic acid molecule encoding CAL linked in frame to a nucleic acid molecule encoding an ecdysone receptor ligand binding domain or a glucocorticoid receptor ligand binding domain. In addition, there is provided a transgenic seed plant containing a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding LFY linked in frame to a nucleic acid molecule encoding an
15 ecdysone receptor ligand binding domain or a glucocorticoid receptor ligand binding domain.

Any floral meristem identity gene product, as defined herein, is useful in a chimeric protein of the invention. Thus, a nucleic acid molecule encoding *Arabidopsis thaliana* AP1 (SEQ ID NO: 2), *Brassica oleracea* AP1 (SEQ ID NO: 4), *Brassica oleracea* var. *Botrytis* AP1 (SEQ ID NO: 8) or *Zea mays* AP1 (SEQ ID NO: 10), each of which have
20 activity in converting shoot meristem to floral meristem, can be used to construct a nucleic acid molecule encoding a chimeric protein of the invention. Similarly, a nucleic acid molecule encoding, for example, *Arabidopsis thaliana* CAL (SEQ ID NO: 10), *Brassica oleracea* CAL (SEQ ID NO: 12), or a nucleic acid molecule encoding *Arabidopsis thaliana* LFY (SEQ ID NO: 16) is useful when linked in frame to a nucleic acid molecule encoding a
25 ligand binding domain to produce a nucleic acid molecule encoding a ligand-dependent chimeric protein of the invention. Similarly, nucleic acids encoding SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27 can be operably linked to a nucleic acid encoding a ligand binding domain.

A ligand binding domain useful in a chimeric protein of the invention is a
30 domain that, when fused in frame to a heterologous gene product, renders the activity of the fused gene product dependent on cognate ligand such that the activity of the fused gene product is increased in the presence of cognate ligand as compared to its activity in the absence of ligand. Such a ligand binding domain can be a steroid binding domain such as the ligand binding domain of an ecdysone receptor, glucocorticoid receptor, estrogen receptor,

progesterone receptor, androgen receptor, thyroid receptor, vitamin D receptor or retinoic acid receptor. A particularly useful ligand binding domain is the ecdysone receptor ligand binding domain contained within amino acids 329 to 878 of the *Drosophila* ecdysone receptor (SEQ ID NO: 18); Koelle et al., *Cell* 67:59-77 (1991); Thummel, *Cell* 83:871-877 (1995), each of which is incorporated herein by reference) or a glucocorticoid receptor ligand binding domain, encompassed, for example, within amino acids 512 to 795 of the rat glucocorticoid receptor (SEQ ID NO: 20; Miesfeld et al., *Cell* 46:389-399 (1986), which is incorporated herein by reference).

A chimeric protein of the invention containing an ecdysone receptor ligand binding domain has floral meristem identity gene product activity that can be increased in the presence of ecdysone ligand. Similarly, a chimeric protein of the invention containing a glucocorticoid receptor ligand binding domain has floral meristem identity gene product activity that is increased in the presence of glucocorticoid ligand. It is well known that in a chimeric protein containing a heterologous gene product such as adenovirus E1A, c-myc, c-fos, the HIV-1 Rev transactivator, MyoD or maize regulatory factor R fused to the rat glucocorticoid receptor ligand binding domain, activity of the fused heterologous gene product can be increased by glucocorticoid ligand (Eilers et al., *Nature* 340:66 (1989); Superti-Furga et al., *Proc. Natl. Acad. Sci., U.S.A.* 88:5114 (1991); Hope et al., *Proc. Natl. Acad. Sci., U.S.A.* 87:7787 (1990); Hollenberg et al., *Proc. Natl. Acad. Sci., U.S.A.* 90:8028 (1993), each of which is incorporated herein by reference).

A nucleic acid molecule encoding a chimeric protein of the invention can be introduced into a seed plant where, under appropriate conditions, the chimeric protein is expressed. In such a transgenic seed plant, floral meristem identity gene product activity can be increased by contacting the transgenic seed plant with cognate ligand. For example, activity of a heterologous protein fused to a rat glucocorticoid receptor ligand binding domain (amino acids 512 to 795) expressed under the control of the constitutive cauliflower mosaic virus 35S promoter in *Arabidopsis* was low in the absence of glucocorticoid ligand; whereas, upon contacting the transformed plants with a synthetic glucocorticoid, dexamethasone, activity of the protein was increased greatly (Lloyd et al., *Science* 266:436-439 (1994), which is incorporated herein by reference). As disclosed herein, a ligand binding domain fused to a floral meristem identity gene product renders the activity of a fused floral meristem identity gene product ligand-dependent such that, upon contacting the transgenic seed plant with cognate ligand, floral meristem identity gene product activity is increased.

Methods for constructing a nucleic acid molecule encoding a chimeric protein of the invention are routine and well known in the art (Sambrook et al., *supra*, 1989). Methods of constructing, for example, a nucleic acid encoding an AP1-glucocorticoid receptor ligand binding domain chimeric protein are described in Example IV of WO 97/46078. For example, the skilled artisan recognizes that a stop codon encoded by the nucleic acid molecule must be removed and that the two nucleic acid molecules must be linked in frame such that the reading frame of the 3' nucleic acid molecule coding sequence is preserved. Methods of transforming a seed plant such as an angiosperm or gymnosperm with a nucleic acid molecule are disclosed above and well known in the art (see Examples I, II and III of WO 97/46078; see, also, Mohoney et al., U.S. Patent Number 5,463,174, and Barry et al., U.S. patent number 5,463,175, each of which is incorporated herein by reference).

As used herein, the term "linked in frame," when used in reference to two nucleic acid molecules that make up a nucleic acid molecule encoding a chimeric protein, means that the two nucleic acid molecules are linked in the correct reading frame such that, under appropriate conditions, a full-length chimeric protein is expressed. In particular, a 5' nucleic acid molecule, which encodes the amino-terminal portion of the chimeric protein, must be linked to a 3' nucleic acid molecule, which encodes the carboxyl-terminal portion of the chimeric protein, such that the carboxyl-terminal portion of the chimeric protein is translated in the correct reading frame. One skilled in the art would recognize that a nucleic acid molecule encoding a chimeric protein of the invention can comprise, for example, a 5' nucleic acid molecule encoding a floral meristem identity gene product linked in frame to a 3' nucleic acid molecule encoding a ligand binding domain or can comprise a 5' nucleic acid molecule encoding a ligand binding domain linked in frame to a 3' nucleic acid molecule encoding a floral meristem identity gene product. Preferably, a nucleic acid molecule encoding a chimeric protein of the invention comprises a 5' nucleic acid molecule encoding a floral meristem identity gene product linked in frame to a 3' nucleic acid molecule encoding a ligand binding domain.

In a transgenic angiosperm containing a chimeric protein of the invention, conversion of shoot meristem to floral meristem can be induced by contacting the transgenic angiosperm with a cognate ligand that is absorbed by the angiosperm and binds the chimeric protein within its ligand binding domain. Thus, the present invention provides a method of converting shoot meristem to floral meristem in an angiosperm by introducing into the angiosperm a nucleic acid molecule encoding a chimeric protein to produce a transgenic angiosperm, where, under appropriate conditions, the chimeric protein containing a floral

meristem identity gene product fused to a ligand binding domain is expressed; and contacting the transgenic angiosperm with cognate ligand, where, upon binding of the cognate ligand to the ligand binding domain, floral meristem identity gene product activity is increased, thereby converting shoot meristem to floral meristem in the transgenic angiosperm.

5 The present invention provides, for example, a method of converting shoot meristem to floral meristem in an angiosperm by introducing into the angiosperm a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, AGL27, AP1, CAL or LFY linked in frame to a nucleic acid molecule encoding an ecdysone receptor ligand binding domain, to
10 produce a transgenic angiosperm, where, under appropriate conditions, the chimeric protein is expressed; and contacting the transgenic angiosperm with ecdysone ligand, where, upon binding of the ecdysone ligand to the ecdysone receptor ligand binding domain, floral meristem identity gene product activity is increased, thereby converting shoot meristem to floral meristem in the transgenic angiosperm. Similarly, the invention provides, for example,
15 a method of converting shoot meristem to floral meristem in an angiosperm by introducing into the angiosperm a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, AGL27, AP1, CAL or LFY linked in frame to a nucleic acid molecule encoding a glucocorticoid receptor ligand binding domain, to produce a transgenic angiosperm, where, under appropriate
20 conditions, the chimeric protein is expressed; and contacting the transgenic angiosperm with glucocorticoid ligand, where, upon binding of the glucocorticoid ligand to the glucocorticoid receptor ligand binding domain, floral meristem identity gene product activity is increased, thereby converting shoot meristem to floral meristem in the transgenic angiosperm.

 In addition, the invention provides a method of promoting early reproductive
25 development in a seed plant by introducing into the seed plant a nucleic acid molecule encoding a chimeric protein of the invention to produce a transgenic seed plant, where, under appropriate conditions, the chimeric protein containing a floral meristem identity gene product fused to a ligand binding domain is expressed; and contacting the transgenic seed plant with cognate ligand, where, upon binding of the cognate ligand to the ligand binding
30 domain, floral meristem identity gene product activity is increased, thereby promoting early reproductive development in the transgenic seed plant. The methods of the invention can be practiced with numerous seed plant varieties. The seed plant can be, for example, an angiosperm such as a cereal plant, leguminous plant, hardwood tree or coffee plant, or can be a gymnosperm such as a pine, fir, spruce or redwood tree.

There is provided, for example, a method of promoting early reproductive development in a seed plant by introducing into the seed plant a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding a floral meristem identity gene product linked in frame to a nucleic acid molecule encoding an ecdysone receptor ligand binding domain, to produce a transgenic seed plant, where, under appropriate conditions, the chimeric protein is expressed; and contacting the transgenic seed plant with ecdysone ligand, where, upon binding of the ecdysone ligand to the ecdysone receptor ligand binding domain, floral meristem identity gene product activity is increased, thereby promoting early reproductive development in the transgenic seed plant. Similarly, the invention provides, for example, a method of promoting early reproductive development in a seed plant by introducing into the seed plant a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding AP1, CAL, LFY, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27 linked in frame to a nucleic acid molecule encoding a glucocorticoid receptor ligand binding domain, to produce a transgenic seed plant, where, under appropriate conditions, the chimeric protein is expressed; and contacting the transgenic seed plant with glucocorticoid ligand, where, upon binding of the glucocorticoid ligand to the glucocorticoid receptor ligand binding domain, floral meristem identity gene product activity is increased, thereby promoting early reproductive development in the transgenic seed plant.

As used herein, the term "ligand" means a naturally occurring or synthetic chemical or biological molecule such as a simple or complex organic molecule, a peptide, a protein or an oligonucleotide that specifically binds a ligand binding domain. In the methods of the present invention, a ligand can be used alone or in solution or can be used in conjunction with an acceptable carrier that can serve to stabilize the ligand or promote absorption of the ligand by a seed plant. If desired, a transgenic seed plant of the invention can be contacted with a ligand for increasing floral meristem identity gene product activity in combination with an unrelated molecule such as a plant nutrient, pesticide or insecticide. When used in reference to a particular ligand binding domain, the term "cognate ligand" means a ligand that, under suitable conditions, specifically binds the particular ligand binding domain.

One skilled in the art readily can determine the optimum concentration of cognate ligand needed to bind a ligand binding domain and increase floral meristem identity gene product activity in a transgenic seed plant of the invention. Generally, a concentration of about 1 nM to 10 μ M cognate ligand is useful for increasing floral meristem identity gene

product activity in a transgenic seed plant expressing a chimeric protein of the invention. Preferably, a concentration of about 100 nM to 1 μ M cognate ligand is useful for increasing floral meristem identity gene product activity in a transgenic seed plant containing a chimeric protein of the invention (see, for example, Christopherson et al., *Proc. Natl. Acad. Sci. USA* 89:6314-6318 (1992), which is incorporated herein by reference; also, see Lloyd et al., *supra*, 1994). For example, a concentration of about 100 nM to 1 μ M dexamethasone can be useful for increasing floral meristem identity gene product activity in a transgenic seed plant of the invention containing a nucleic acid molecule encoding a chimeric protein, which comprises a nucleic acid molecule encoding a floral meristem identity gene product, such as AP1, CAL, LFY, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27 linked in frame to a nucleic acid molecule encoding a glucocorticoid receptor ligand binding domain.

As discussed above, a transgenic seed plant of the invention, such as a transgenic seed plant expressing a chimeric protein of the invention, can be contacted in a variety of manners. A transgenic seed plant can be contacted with cognate ligand, for example, by spraying the seed plant with a gaseous ligand or with solution such as an aqueous solution containing the appropriate ligand; or by adding the cognate ligand to the water supply of a seed plant grown using irrigation or grown hydroponically; or by adding the cognate ligand to the soil or other solid nutrient medium in which a seed plant is grown, whereby the cognate ligand is absorbed into the seed plant to increase floral meristem identity gene product activity. A transgenic seed plant expressing a chimeric protein of the invention also can be contacted with a cognate ligand in aerosol form. In addition, a transgenic seed plant can be contacted with cognate ligand by injecting the seed plant or by immersing the seed plant in a solution containing the cognate ligand.

A transgenic seed plant expressing a chimeric protein of the invention can be contacted individually with cognate ligand, or a group of transgenic seed plants can be contacted *en masse* to increase floral meristem gene product activity synchronously in all seed plants of the group. Furthermore, a variety of means can be used to contact a transgenic seed plant of the invention with cognate ligand to increase floral meristem identity gene product activity. A transgenic seed plant can be contacted with cognate ligand using, for example, a hand held spraying apparatus; conventional yard "sprinkler"; mechanical spraying system, such as an overhead spraying system in a green house; traditional farm machinery, or "crop dusting." As discussed above in regard to the application of inducing agents, the methods of the invention can be practiced using these and other manual or mechanical means to contact a transgenic seed plant with single or multiple applications of cognate ligand.

IX. Nucleic Acid Molecules of the Invention

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art.

5 Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

15 The isolation of nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as a floral organ, and a cDNA library which contains the gene transcript of interest is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which genes of the invention or
20 homologs are expressed.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned nucleic acid disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against an polypeptide can be used to screen an
25 mRNA expression library.

Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the nucleic acid of the invention directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other
30 *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR see *PCR Protocols: A Guide to Methods and*

Applications. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990). Appropriate primers and probes for identifying sequences from plant tissues are generated from comparisons of the sequences provided here with other related genes.

The present invention also provides novel substantially purified nucleic acid molecules encoding gene products including AP1, CAL, LFY, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, and AGL27. For example, the invention provides a substantially purified nucleic acid molecule encoding *Brassica oleracea* AP1 having the amino acid sequence SEQ ID NO:4; a substantially purified nucleic acid molecule encoding *Brassica oleracea* var. *botrytis* AP1 having the amino acid sequence SEQ ID NO:6; or a substantially purified nucleic acid molecule encoding *Zea mays* AP1 having the amino acid sequence SEQ ID NO: 8. In addition, the invention provides a substantially purified nucleic acid molecule that encodes a *Brassica oleracea* AP1, *Brassica oleracea* var. *botrytis* AP1 or *Zea mays* AP1 and that contains additional 5' or 3' noncoding sequence. For example, a substantially purified nucleic acid molecule having a nucleotide sequence such as SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7 is provided.

The invention also provides a substantially purified nucleic acid molecule encoding a CALIFLOWER gene product such as *Arabidopsis thaliana* CAL (SEQ ID NO: 10) or *Brassica oleracea* CAL (SEQ ID NO: 12). The invention also provides nucleic acid molecules encoding SEP1 (SEQ ID NO:28), SEP2 (SEQ ID NO:30), SEP3 (SEQ ID NO:32), AGL20 (SEQ ID NO:34), AGL22 (SEQ ID NO:36), AGL24 (SEQ ID NO:38) or AGL27 (SEQ ID NO:40).

As used herein in reference to a particular nucleic acid molecule or gene product, the term "substantially purified" means that the particular nucleic acid molecule or gene product is in a form that is relatively free from contaminating lipids, unrelated gene products, unrelated nucleic acids or other cellular material normally associated with the particular nucleic acid molecule or gene product in a cell.

The present invention also provides a nucleotide sequence having at least ten contiguous nucleotides of a nucleic acid molecule encoding any of the above-referenced gene products, including *Brassica oleracea* AP1, *Brassica oleracea* var. *botrytis* AP1 or *Zea mays* AP1, provided that said nucleotide sequence is not present in a nucleic acid molecule encoding a MADS domain containing protein. For example, such a nucleotide sequence can have at least ten contiguous nucleotides of a nucleic acid molecule encoding an AP1 gene product having the amino acid sequence of SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8.

A nucleotide sequence of the invention can have, for example, at least ten contiguous nucleotides of the nucleic acid sequence of SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7.

As used herein, the term "contiguous," as used in reference to the nucleotides of a nucleic acid molecule means that the nucleotides of the nucleic acid molecule follow continuously in sequence. Thus, a nucleotide sequence of the invention has at least ten contiguous nucleotides of one of the recited nucleic acid molecules without any extraneous intervening nucleotides.

Explicitly excluded from a nucleotide sequence of the present invention is a nucleotide sequence having at least ten contiguous nucleotides that is present in a nucleic acid molecule encoding a MADS domain containing protein. MADS domain containing proteins are well known in the art as described in Purugganan et al., *supra*, 1995.

In general, a nucleotide sequence of the invention can range in size from about 10 nucleotides to the full-length of a cDNA. Such a nucleotide sequence can be chemically synthesized, using routine methods or can be purchased from a commercial source. In addition, such a nucleotide sequence can be obtained by enzymatic methods such as random priming methods, polymerase chain reaction (PCR) methods or by standard restriction endonuclease digestion, followed by denaturation (Sambrook et al., *supra*, 1989).

A nucleotide sequence of the invention can be useful, for example, as a primer for PCR (Innis et al. (ed.) *PCR Protocols: A Guide to Methods and Applications*, San Diego, CA: Academic Press, Inc. (1990)). Such a nucleotide sequence generally contains from about 10 to about 50 nucleotides.

A nucleotide sequence of the invention also can be useful in screening a cDNA or genomic library to obtain a related nucleotide sequence. For example, a cDNA library that is prepared from rice or wheat can be screened with a nucleotide sequence having at least ten contiguous nucleotides of the nucleic acid molecule encoding *Zea mays AP1* (SEQ ID NO: 7) in order to isolate a rice or wheat ortholog of *AP1*. Generally, a nucleotide sequence useful for screening a cDNA or genomic library contains at least about 14 to 16 contiguous nucleotides depending, for example, on the hybridization conditions to be used. A nucleotide sequence containing at least 18 to 20 nucleotides, or containing at least 21 to 25 nucleotides, also can be useful.

A nucleotide sequence having at least ten contiguous nucleotides of a nucleic acid molecule encoding *Zea mays AP1* (SEQ ID NO: 7) also can be used to screen a *Zea mays* cDNA library to isolate a sequence that is related to but distinct from *AP1*. Similarly, a nucleotide sequence having at least ten contiguous nucleotides of a nucleic acid molecule

encoding *Brassica oleracea* AP1 (SEQ ID NO: 3) or a nucleotide sequence having at least ten contiguous nucleotides of a nucleic acid molecule encoding *Brassica oleracea* var.

botrytis AP1 (SEQ ID NO: 5) can be used to screen a *Brassica oleracea* or *Brassica oleracea* var. *botrytis* cDNA library to isolate a novel sequence that is related to but distinct from AP1.

5 Other gene orthologs, such as of SEP1, SEP2, SEP3, AGL20, AGL22, AGL24 or AGL27 can be isolated by similar methods. In addition, a nucleotide sequence of the invention can be useful in analyzing RNA levels or patterns of expression, as by northern blotting or by *in situ* hybridization to a tissue section. Such a nucleotide sequence also can be used in Southern blot analysis to evaluate gene structure and identify the presence of related gene sequences.

10 The invention also provides a vector containing a nucleic acid molecule as described above, e.g., encoding a *Brassica oleracea* AP1 gene product, *Brassica oleracea* var. *botrytis* AP1 gene product or *Zea mays* AP1 gene product. A vector can be a cloning vector or an expression vector and provides a means to transfer an exogenous nucleic acid molecule into a host cell, which can be a prokaryotic or eukaryotic cell. Such vectors are well known and include plasmids, phage vectors and viral vectors. Various vectors and methods for introducing such vectors into a cell are described, for example, by Sambrook et al., *supra*, 1989, and by Glick and Thompson, *supra*, 1993).

20 The invention further provides a method of producing one of the above-described gene products by expressing a nucleic acid molecule encoding the gene product (e.g., AP1, CAL, SEP1, SEP2, SEP3, AGL20, AGL22, AGL24, or AGL27). Thus, for example, a *Brassica oleracea* AP1 gene product can be produced according to a method of the invention by expressing a nucleic acid molecule having the amino acid sequence of SEQ ID NO: 4 or by expressing a nucleic acid molecule having the nucleic acid sequence of SEQ ID NO: 3. Similarly, a *Brassica oleracea* var. *botrytis* AP1 gene product can be produced according to a method of the invention by expressing a nucleic acid molecule having the amino acid sequence of SEQ ID NO: 6 or by expressing a nucleic acid molecule having the nucleic acid sequence of SEQ ID NO: 5. A *Zea mays* AP1 gene product can be produced by expressing a nucleic acid molecule having the amino acid sequence of SEQ ID NO: 8 or by expressing a nucleic acid molecule having the nucleic acid sequence of SEQ ID NO: 7.

30 The invention also provides a substantially purified AP1 gene product, such as a substantially purified gene product of the invention such as a *Brassica oleracea* AP1 gene product having amino acid sequence SEQ ID NO: 4; a substantially purified *Brassica oleracea* var. *botrytis* AP1 gene product having amino acid sequence SEQ ID NO: 6; or a substantially purified *Zea mays* AP1 gene product having amino acid sequence SEQ ID NO:

8. As used herein, the term "gene product" is used in its broadest sense and includes proteins, polypeptides and peptides, which are related in that each consists of a sequence of amino acids joined by peptide bonds. For convenience, the terms "gene product," "protein" and "polypeptide" are used interchangeably. While no specific attempt is made to distinguish the size limitations of a protein and a peptide, one skilled in the art would understand that proteins generally consist of at least about 50 to 100 amino acids and that peptides generally consist of at least two amino acids up to a few dozen amino acids. The term gene product as used herein includes any such amino acid sequence.

An active fragment of a floral meristem identity gene product also can be useful in the methods of the invention. As used herein, the term "active fragment," means a polypeptide portion of a floral meristem identity gene product that can convert shoot meristem to floral meristem in an angiosperm. An active fragment of an AP1 gene product can consist, for example, of an amino acid sequence that is derived from SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6 or SEQ ID NO: 8 and has activity in converting shoot meristem to floral meristem in an angiosperm. An active fragment can be, for example, an amino terminal, carboxyl terminal or internal fragment of *Zea mays* AP1 (SEQ ID NO: 8) that has activity in converting shoot meristem to floral meristem in an angiosperm. The skilled artisan will recognize that an active fragment of a floral meristem identity gene product, as defined herein, can be useful in the methods of the invention for converting shoot meristem to floral meristem in an angiosperm, for producing early reproductive development in a seed plant, or for producing reproductive sterility in a seed plant.

Such an active fragment can be produced using well known recombinant DNA methods (Sambrook et al., *supra*, 1989). Similarly, an active fragment can be, for example, an amino terminal, carboxyl terminal or internal fragment of *Arabidopsis thaliana* CAL (SEQ ID NO: 10) or *Brassica oleracea* CAL (SEQ ID NO: 12) that has activity, for example, in converting shoot meristem to floral meristem in an angiosperm. The product of the *BobCAL* gene (SEQ ID NO: 24), which is truncated at amino acid 150, lacks activity in converting shoot meristem to floral meristem and, therefore, is an example of a polypeptide portion of a CAL floral meristem identity gene product that is not an "active fragment" of a floral meristem identity gene product.

An active fragment of a floral meristem identity gene product, which can convert shoot meristem to floral meristem in an angiosperm, can be identified using the methods described in WO 97/46078. Briefly, an angiosperm such as *Arabidopsis* can be transformed with a nucleic acid molecule encoding a portion of a floral meristem identity

gene product in order to determine whether the portion can convert shoot meristem to floral meristem and, therefore, is an active fragment of a floral meristem identity gene product.

The invention also provides an expression vector containing a nucleic acid molecule encoding a floral meristem identity gene product such as SEP3, AGL20, AGL22, AGL24, AGL27, AP1, CAL or LFY operably linked to a heterologous regulatory element. Expression vectors are well known in the art and provide a means to transfer and express an exogenous nucleic acid molecule into a host cell. Thus, an expression vector contains, for example, transcription start and stop sites such as a TATA sequence and a poly-A signal sequence, as well as a translation start site such as a ribosome binding site and a stop codon, if not present in the coding sequence.

As used herein, the term "heterologous regulatory element" means a regulatory element derived from a different gene than the gene encoding the floral meristem identity gene product to which it is operably linked. A vector containing a floral meristem identity gene, however, contains a nucleic acid molecule encoding a floral meristem identity gene product operably linked to a homologous regulatory element. Such a vector does not contain a nucleic acid molecule encoding a floral meristem identity gene product operably linked to a heterologous regulatory element and, thus, is not an expression vector of the invention.

The invention further provides a plant expression vector containing a floral meristem identity gene product operably linked to a heterologous regulatory element. For example, a plant expression vector containing a nucleic acid molecule encoding an AP1 gene product having at least about 70 percent amino acid identity with an amino acid sequence of *Arabidopsis thaliana* AP1 (SEQ ID NO: 2) in the region from amino acid 1 to amino acid 163 or with the amino acid sequence of *Zea mays* AP1 (SEQ ID NO: 8) in the region from amino acid 1 to amino acid 163 is provided. A plant expression vector containing a floral meristem identity gene product operably linked to a constitutive regulatory element, such as the cauliflower mosaic virus 35S promoter, is provided. In addition, a plant expression vector containing a floral meristem identity gene product operably linked to an inducible regulatory element is provided.

A useful plant expression vector can contain a constitutive regulatory element for expression of an exogenous nucleic acid molecule in all or most tissues of a seed plant. The use of a constitutive regulatory element can be particularly advantageous because expression from the element is relatively independent of developmentally regulated or tissue-specific factors. For example, the cauliflower mosaic virus 35S promoter (CaMV 35S) is a well-characterized constitutive regulatory element that produces a high level of

expression in all plant tissues (Odell et al., *Nature* 313:810-812 (1985), which is incorporated herein by reference). Furthermore, the CaMV 35S promoter can be particularly useful due to its activity in numerous different seed plant species (Benfey and Chua, *Science* 250:959-966 (1990), which is incorporated herein by reference; Odell et al., *supra*, 1985). Other

5 constitutive regulatory elements useful for expression in a seed plant include, for example, the cauliflower mosaic virus 19S promoter; the Figwort mosaic virus promoter (Singer et al., *Plant Mol. Biol.* 14:433 (1990), which is incorporated herein by reference); and the nopaline synthase (*nos*) gene promoter (An, *Plant Physiol.* 81:86 (1986), which is incorporated herein by reference).

10 In addition, an expression vector of the invention can contain a regulated gene regulatory element such as a promoter or enhancer element. A particularly useful regulated promoter is a tissue-specific promoter such as the shoot meristem-specific *CDC2* promoter (Hemerly et al., *Plant Cell* 5:1711-1723 (1993), which is incorporated herein by reference), or the *AGL8* promoter, which is active in the apical shoot meristem immediately after the

15 transition to flowering (Mandel and Yanofsky, *supra*, 1995). The promoter of the *SHOOTMERISTEMLESS* gene, which is expressed exclusively in the shoot meristem beginning within an embryo and throughout the angiosperm life cycle, also can be a particularly useful tissue-specific gene regulatory element (see Long et al., *Nature* 379:66-69 (1996), which is incorporated herein by reference).

20 An appropriate regulatory element such as a promoter is selected depending on the desired pattern or level of expression of a nucleic acid molecule linked thereto. For example, a constitutive promoter, which is active in all tissues, would be appropriate if expression of a gene product in all plant tissues is desired. In addition, a developmentally regulated or tissue-specific regulatory element can be useful to direct floral meristem identity

25 gene expression to specific tissues, for example. As discussed above, inducible expression also can be particularly useful to manipulate the timing of gene expression such that, for example, a population of transgenic seed plants of the invention that contain an expression vector comprising a floral meristem identity gene linked to an inducible regulatory element can undergo early reproductive development at essentially the same time. Selecting the time

30 of reproductive development can be useful, for example, in manipulating the time of crop harvest.

Using nucleic acid molecules encoding gene products provided herein, the skilled artisan can isolate, if desired, novel orthologs. For example, one would choose a region of *API* that is highly conserved among known *API* sequences such as a region that is

highly conserved between *Arabidopsis AP1* (SEQ ID NO: 1) and *Zea mays AP1* (GenBank accession number L46400; SEQ ID NO: 7) to screen a cDNA or genomic library of interest for a novel AP1 ortholog. One can use a full-length *Arabidopsis AP1* (SEQ ID NO: 1), for example, to isolate a novel ortholog of *AP1* (see, e.g., Example V of WO 97/46078). If

desired, the region encoding the MADS domain, which is common to a number of genes, can be excluded, from the sequence used as a probe. Similarly, the skilled artisan knows that a nucleic acid molecule encoding a full-length *CAL* cDNA such as *Arabidopsis CAL* (SEQ ID NO: 9) or *Brassica oleracea CAL* (SEQ ID NO: 11) can be useful in isolating a novel *CAL* ortholog.

For example, the *Arabidopsis AP1* cDNA (SEQ ID NO: 1) can be used as a probe to identify and isolate a novel *AP1* ortholog. Using a nucleotide sequence derived from a conserved region of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 or SEQ ID NO: 7, for example, a nucleic acid molecule encoding a novel *AP1* ortholog can be isolated from other plant species. Using methods such as those described by Purugganan et al., *supra*, 1995, one can readily confirm that the newly isolated molecule is an *AP1* ortholog. Thus, a nucleic acid molecule encoding an AP1 gene product, which has at least about 70 percent amino acid identity with the amino acid sequence of SEQ ID NO: 2 (*Arabidopsis AP1*) in the region from amino acid 1 to amino acid 163 or with the amino acid sequence of SEQ ID NO: 8 (*Zea mays AP1*) in the region from amino acid 1 to amino acid 163 can be isolated and identified using well known methods.

Similarly, in order to isolate an ortholog of *CAL*, one can choose a region of *CAL* that is highly conserved among known *CAL* cDNAs, such as a region conserved between *Arabidopsis CAL* (SEQ ID NO: 9) and *Brassica oleracea CAL* (SEQ ID NO: 11). The *Arabidopsis CAL* cDNA (SEQ ID NO: 9) or *Brassica oleracea CAL* cDNA (SEQ ID NO: 11), or a nucleotide fragment thereof, can be used to identify and isolate a novel *CAL* ortholog using methods such as those described in Example V of WO 97/46078. In order to identify related MADS domain genes, a nucleotide sequence derived from the MADS domain of *AP1* or *CAL*, for example, can be useful to isolate a related gene sequence encoding this DNA-binding motif.

Hybridization conditions for isolating a gene ortholog, for example, are relatively stringent such that non-specific hybridization is minimized. Appropriate hybridization conditions can be determined empirically, or can be estimated based, for example, on the relative G+C content of the probe and the number of mismatches between the probe and target sequence, if known. Hybridization conditions can be adjusted as desired

by varying, for example, the temperature of hybridizing or the salt concentration (Sambrook, *supra*, 1989).

The invention also provides a kit for converting shoot meristem to floral meristem in an angiosperm, which contains a plant expression vector having a nucleic acid molecule encoding a floral meristem identity gene product. A kit for promoting early reproductive development in a seed plant, which contains a plant expression vector having a nucleic acid molecule encoding a floral meristem identity gene product, also is provided. If desired, such kits can contain appropriate reagents to facilitate high efficiency transformation of a seed plant with a plant expression vector of the invention. Furthermore, if desired, a control vector lacking a floral meristem identity gene can be included in the kits to determine, for example, the efficiency of transformation.

The following example is offered by way of example, not limitation.

EXAMPLES

Example 1

This example shows the identification of proteins that interact with CAL.

Proteins that interact with CAL

Yeast two-hybrid screens were performed to identify candidate genes whose products interact with AP1 and CAL. The two-hybrid library screens were performed in the YPB2 strain [*MATa ara3 his3 ade2 lys2 trp1 leu2, 112 can^r gal4 gal80 LYS2::GAL1-HIS3, URA3::(GAL1 UAS17mers)-lacZ*]. Yeast were transformed using a modified version of the lithium acetate method of Schiestl and Gietz, *Curr. Genet.* 16, 339-346 (1989).

The two-hybrid cDNA expression library was constructed in the pBI771 (prey) vector using tissue of whole plants at different stages. The bait constructs were prepared by inserting the intact CAL coding region and a truncated form of AP1 into the pBI-880 vector (a variant of pPC62 described in Chevray and Nathans *Proc. Natl. Acad. Sci. USA* 5789-5793 (1992); Kohalmi *et al.*, *Plant. Mol. Biol. Man.* M1, 1-30 (1998)) by inserting the corresponding coding region in-frame at the 3' end of the GAL4 (1-147) sequence contained in the centromere *LEU2* plasmid. These baits tested negative for the ability to activate transcription of both reporters, alone as well as in combination with each the prey vector and an inert control prey, the *Arabidopsis* cruciferin seed storage protein.

SEP3K, SOC1K, SVPK, AGL24K and SOC1KC/2 were generated by polymerase chain reaction (PCR) from the relevant cDNAs using oligos with the appropriate restriction site for posterior cloning into pBI771. The following primers were used:

SEP3-5'K: 5'-CCGTCGACCCATGAGCCAGCAGGAGTATCTC-3'

SEP3-3'Kbox: 5'-CCGCGGCCGCCTTACTCTGAAGATCGTT-3'

SOC1-5'K: 5'-CCGTCGACCCATGAAATATGAAGCAGCAAAC-3'

SOC1-3'Kbox: 5'-CCGCGGCCGCCTCCTTTTGCTTGAGCTG-3'

SOC1-C/2: 5'-CCGCGGCCGCCTTTCTTGATTCTTATT-3'

SVP-5'K: 5'-CCGTCGACCCATGAGTGATCACGCCCGAATG-3'

SVP-3'Kbox: 5'-CCGCGGCCGCCTCCCTTTTCTGAAGTTC-3'

AGL24-5'K: 5'-CCGTCGACCCATGCTTGAGAATTGTAACCTC-3'

AGL24-3'Kbox: 5'-CCGCGGCCGCCTCAAGTGAGAAAATTTG-3'

The PCR products were subcloned directly into pCRII (Invitrogen) and then digested with SalI-NotI for next subcloning into pBI-771. All constructs were confirmed by sequencing.

CAL screen: The frequency of clones which activated both the *HIS3* and *lacZ* reporters from the 30°C plates was $1/(1.8 \times 10^6) = 5.6 \times 10^{-7}$. The frequency on the 23°C plates was $22/(1.8 \times 10^6) = 1.2 \times 10^{-5}$.

AP1 screen: 9.2×10^4 total transformants were screened at 23°C and the frequency of clones activating both reporter genes was 1.5×10^{-4} .

The transformants were selected on supplemented synthetic dextrose medium lacking leucine, tryptophan and histidine but containing 5 mM 3-amino-1,2,4-triazole. The colonies growing on this selective medium were assayed for β -galactosidase activity on nitrocellulose filters (Kohalmi *et al.*, *supra*). Plasmid DNA from positive clones was isolated and transformed into *E. coli*.

Using a full-length CAL cDNA as bait, 23 interacting clones were identified, rescued from yeast and transformed into *E. coli*. Sequence analyses showed that they fell into four classes, all previously identified as *AGAMOUS-like* (*AGL*) genes.

The first class, *SEP3*, included four clones, all of which began within the I-region. Because the cDNA library was poly (T) primed, the clones all comprised varying lengths of the 3' end of the gene. *SEP3* is first expressed in the central dome of stage-two floral primordia and is maintained in the inner three whorls of the flower (Mandel and Yanofsky, *Sex. Plant Reprod.* 11, 22-28 (1998)). *SEP3* acts redundantly with *SEP1* and

SEP2 and is necessary for the development of petals, stamens and carpels (Pelaz *et al.*, *Current Biology* 11, 182-184 (2000)).

The second class identified was the *SUPPRESSOR OF CO OVEREXPRESSION 1 (SOC1)* gene and included seven clones. The starting point of these clones varied. One clone began with the ATG start codon, another started near the end of the MADS-box, and the remaining clones started at 5' ends of the I-region. *SOC1* is expressed in the inflorescence meristem, as well as in the two inner whorls of the flower beginning in late stage-two and it is involved in promoting flowering (Samach *et al.*, *Science* 288, 1613-1616 (2000)).

The third class was the *SHORT VEGETATIVE PHASE (SVP)* gene, and included four clones. Of the clones from this screen, one started in the MADS-box, and three began in the I-region. *SVP* was identified as an *Arabidopsis* expressed sequence tag with homology to the MADS-box family (Alvarez-Buylla *et al.*, *Plant J.* 24, 457-466 (2000)), and it was also cloned by (Hartman *et al.*, 2000) through transposon tagging. *SVP* is a repressor of flowering and is expressed in young leaves and throughout the shoot apical meristem during vegetative development. After the transition to flowering, it is expressed in young flower primordia until stage 3 (Hartman *et al.*, *Plant J.* 21, 351-360 (2000)).

The last eight clones were identified as *AGL24*. One of these clones began within the MADS-box and three within the I-region. In addition, the 5' ends of four clones lie in the first third of the K-box, representing the shortest clones isolated in the screen. *AGL24* was first identified in a previous yeast two-hybrid screen as a clone which interacts with *AG* (Alvarez-Buylla *et al.*, *Proc. Natl. Acad. Sci. USA* 97, 5328-5333 (2000)). *AGL24* is expressed in inflorescences and young floral primordia.

To confirm the specificity of the observed interactions, the longest and shortest clone of each class was transformed back into a yeast strain that contained either the CAL bait, the bait vector, or an inert control bait, cruciferin. The strains containing the CAL bait tested positive for both β -Gal activity and *HIS* prototrophy. The strains containing the bait vector or the cruciferin bait were negative in both assays, as they were not able to grow on plates lacking histidine and the yeast colonies were completely white in the β -Gal assay.

AP1 forms dimers in yeast with CAL interactors

The structural and functional similarities between CAL and AP1 suggested that they may interact with an overlapping set of proteins. In order to explore this possibility,

we constructed an AP1 bait by inserting the intact AP1 coding region into the pBI-880 vector. As in the Finley and Brent system, the full-length AP1 bait activated transcription independently. To overcome this problem, a deletion construct was made encoding residues 1-196 of AP1 (AP1 Δ 1), thus eliminating the putative *trans*-activating C-terminus. In contrast to the full-length AP1 clone, the deletion derivative did not activate the reporter on its own. The longest clone of each class was transformed into yeast in combination with the AP1 deletion bait. In every case, both of the reporters were strongly activated, suggesting that all four CAL-interacting proteins also interact with AP1.

Domain for protein-protein interactions

Previous studies have shown that the MADS-domain and I regions may be important for homodimer formation by AG and by AP1 (Krizek and Meyerowitz, 1996; Mizukami *et al.*, 1996; Riechmann *et al.*, 1996) and that the I region and K-domain are needed for the formation of AP3/PI heterodimers (Krizek and Meyerowitz, *Proc. Natl. Acad. Sci. USA* 93, 4063-4070 (1996); Riechmann *et al.*, *Proc. Natl. Acad. Sci. USA* 93, 4793-4798 (1996)). In addition, the K-domain of AG is sufficient to promote interactions with SEP1, SEP2, SEP3 and AGL6 in yeast (Fan *et al.*, *Plant J.* 11, 999-1010 (1997)). Since many of the CAL- and AP1-interacting clones isolated as part of our study lacked the MADS-domain and I regions, we tested if the K-domain itself was sufficient to promote the observed interactions. First, we subcloned the K-box regions of *SEP3*, *SOC1*, *SVP* and *AGL24* into the prey vector, and tested their ability to activate the reporter using either the empty bait or the cruciferin gene cloned into the bait plasmid. As expected, these K-box regions did not activate the reporter. In contrast, when these K-box prey constructs were introduced into yeast strains that contained each of the CAL or AP1 bait plasmids, reporter activity significantly above background levels was consistently observed. Furthermore, the addition of approximately half of the C-terminal domain of the *SOC1* protein was sufficient to greatly strengthen the interaction, similar to what has previously been shown to occur for AG and its interactors (Fan *et al.*, *supra*). Taken together, these studies suggest that the ability of CAL and AP1 to interact with *SEP3*, *SOC1*, *SVP*, and *AGL24* is largely mediated by the K-domain. However, other protein domains appear to enhance these interactions since the level of reporter gene activation is higher when larger constructs are used.

Example 2

This example shows the indentifications of proteins that interact with AP1.

Proteins that interact with AP1

5 In order to find additional proteins that could interact with AP1, the library was screened with the truncated AP1 bait (1-196), and 13 clones that tested positive for β -Gal activity were characterized. As expected, we found three clones of *AGL20* (also known as *SOCI*), five clones of *AGL22* (also known as *SVP*), and one clone of *AGL24*.

10 In addition we found one clone of a new MADS box gene designated *AGL27* (Alvarez-Buylla *et al.*, *supra*), two different clones encoding a putative RNA binding protein (GI 10178188), and one clone encoding a novel protein (GI 3157943). We determined that these three newly isolated genes have overlapping expression patterns with that of *AP1*, consistent with the idea that they may interact with AP1 *in planta*.

15 To confirm the specificity of these interactions, the longest clone of each class was transformed back into yeast with the AP1 bait, the bait vector, and an inert control bait, cruciferin. The strains containing the AP1 bait tested positive for both β -Gal activity and *HIS* prototrophy. The strains containing the bait vector or the cruciferin bait were negative in both assays. We then tested if the three new AP1-interacting clones could also interact with CAL, since they had not been isolated in the CAL library screen. However, *AGL27*, the
20 RNA binding protein, and the novel protein were unable to interact with CAL in yeast.

Example 3

This example demonstrates the characterization of *sep3* mutants.

25 *sep3* mutants resemble intermediate alleles of *API*

As a start toward determining if the observed interactions in yeast reflect functional interactions *in vivo*, we characterized loss- and gain-of-function alleles of *SEP3*. If some of the activities of AP1 require an interaction with SEP3, then mutations in *SEP3* might be expected to resemble mutant alleles of *API*. We recently identified two independently
30 derived En-1 transposon insertion alleles of *SEP3* and have described the phenotype of *sep1 sep2 sep3* triple mutants in which the three inner whorls of organs become sepaloid (Pelaz *et al.*, *supra*).

The flowers of *sep3-1* and *sep3-2* single mutant plants have petals that are partially transformed into sepals, and infrequently, axillary flowers develop at the base of the first-whorl sepals. When examined by scanning electronic microscopy (SEM), the abaxial cells of these transformed petals resemble cells that are a mixture of abaxial wild type sepal and abaxial wild type petal cells. The abaxial side of the wild type sepals have rectangular cells of varying size, some of which are very long, reaching 300 μm in length. These long cells can be more than ten times the length of the smallest sepal cells. Numerous stomata are visible throughout wild type sepals but are never found on wild-type petals. Cells on the abaxial side of wild type petals all have a uniformly small rounded appearance, and are typically about half of the size of the smallest sepal cells. Unlike wild type petals that have rounded cells, the abaxial side of the *sep3* petals consists of rectangular cells, resembling those found on sepals. Although these mutant petal cells are larger than their wild type counterparts, they are still smaller than the wild type sepal cells. Interestingly, several stomata are interspersed on the surface of these petals, further suggesting a partial transformation of these petals into sepals.

Because the *sep3* petal phenotype resembles that observed for intermediate alleles of *ap1*, (Bowman *et al.*, *Development* 119, 721-743 (1993)), we compared second whorl organs of *sep3* mutants to those of intermediate alleles of *ap1*, including *ap1-2*, *ap1-4* and *ap1-6*. The abaxial cells of these *ap1* mutant petals are very similar to those of the *sep3* mutants, and consist of a blend between petal and sepal cells. These *ap1* mutant cells are larger and more elongated than the wild type petal cells but they do not reach the length of the longer wild type sepal cells. As was observed for *sep3* mutants, petals of these intermediate alleles of *ap1* develop several stomata, further indicating the sepal-like identity. The similarities of *sep* and *ap1* mutants are consistent with the idea that some of the activities of *AP1* are compromised in *sep* mutants, consistent with the possible loss of AP1/SEP interactions.

If the interaction between SEP and AP1 is necessary for AP1 activity, then a reduction in *SEP* expression would be predicted to produce some or all of the *ap1*-mutant phenotypes. To test this idea, we generated transgenic antisense lines in which the 5' end of the *SEP3* gene was expressed in the antisense orientation from the double 35S promoter. Two independent transgenic lines (SP70.1 and SP70.2) were tested for reduction in the amount of *SEP3* mRNA accumulation. As expected, the amount of *SEP3* mRNA in these antisense lines was reduced in comparison to the wild type. The resulting lines underexpressing *SEP3*

showed green petals whose cells appeared partially transformed into sepal cells. These plants also occasionally had axillary flowers arising from the base of the first-whorl sepals. These phenotypes are consistent with a reduction in *AP1* activity, as intermediate alleles of *ap1* produce similar phenotypes. This activity reduction does not mean less *AP1* transcription, the levels of mRNA in these antisense lines are comparable to those of wild type flowers. Interestingly, the green-petal phenotype of these *SEP3* antisense lines is more extreme than that observed for *sep3* single mutants, based on the color change, suggesting that the *SEP3* transgene may also have down regulated other closely related genes such as *SEP1* and *SEP2*.

Example 4

This example demonstrates the characterization of plants overexpressing *SEP3*.

Constitutive expression of *SEP3*

Previous studies have demonstrated that constitutive expression of *AP1* (35S::*AP1*) results in plants that flower considerably earlier than wild type plants (Mandel and Yanofsky, *supra*). If some of the activities of *AP1* require an interaction with *SEP3*, as the loss of function studies above would indicate, then it might be expected that constitutive *SEP3* expression would further enhance the 35S::*AP1* early-flowering phenotype. To test this hypothesis and to provide further evidence that *SEP3* interacts with *AP1* *in planta*, we generated 35S::*SEP3* sense lines that express constitutively *SEP3* throughout the plant.

Construction of the 35S::*SEP3* construct was as follows: cDNA was isolated by RT-PCR using the oligos OAM37: 5'-TAGAAACATCATCTTAAAAAT-3' and *SEP3*-5': 5'-CCGGATCCAAAATGGGAAGAGGGAGA-3'. This cDNA was first cloned into pCRII (invitrogen) and then digested with BamHI for insertion into the BamHI site of pCGN18 (which contains 35S promoter) to produce sense lines, and confirmed by sequencing. The cDNA cloned into pCRII was digested with BamHI and BglII, the 363bp band corresponding to the 5' end of the cDNA was cloned in antisense orientation into the BamHI site of pBIN-JIT (plasmid carrying two 35S promoters in tandem). The 35S::*SEP3* sense and antisense constructs were introduced into *Arabidopsis*, ecotype *Columbia*, by vacuum infiltration (Bechtold *et al.*, *C. R. Acad. Sci.* 316, 1194-1199 (1993)) and transgenic plants were selected on Kanamycin plates.

35S::SEP3 transgenic plants are early flowering, and bolt after producing only four or five rosette leaves, in contrast to wild-type plants which bolt after producing approximately ten leaves under these growth conditions. In addition to the early-flowering phenotype, 35S::SEP3 plants have curled rosette leaves as well as two or three very curled cauline leaves, each of which typically subtends a solitary flower. The primary inflorescence usually produces only a few flowers before terminating. Some of the phenotypes caused by ectopic *SEP3* expression are similar to those conferred by ectopic expression of several other MADS-box genes. However ectopic expression of these other genes often produces additional phenotypes, including alterations in flower organ identity and fruit development that are not seen in the 35S::SEP3 plants.

Example 5

This example demonstrates genetic interactions between 35S::SEP3 and 35S::AP1 transgenes.

To provide genetic evidence that SEP3 and AP1 interact, we crossed the 35S::SEP3 transgene into 35S::AP1 plants. Whereas 35S::AP1 plants flower early after producing four to five rosette leaves, 35S::AP1 35S::SEP3 doubly transgenic plants flower after producing only two rosette leaves, often developing a terminal flower directly from the rosette. Occasionally, these plants produce a very short inflorescence with two cauline leaves that subtend solitary flowers, a terminal flower at the apex, and very little internode elongation. The strong enhancement of the early-flowering phenotypes conferred by each single transgene is consistent with the suggestion that AP1 and SEP3 interact *in planta*.

We also used another genetic approach to investigate the interaction between SEP3 and AP1, avoiding the use of two different transgenic lines. We took advantage of the *tfl1* mutant, in which *AP1* is ectopically activated (Bowman *et al.*, *supra*; Gustafson-Brown *et al.*, *Cell* **76**, 131-143 (1994)), producing a phenotype that closely resembles the 35S::AP1 phenotype. As expected, the *tfl* mutation in combination with the 35S::SEP3 transgene produces the same phenotypes as observed for plants carrying both 35S::AP1 and 35S::SEP3 transgenes. These plants flower after forming two rosette leaves and produce abbreviated shoots with very short internodes and a terminal flower.

Example 6

This example demonstrates the flowering time of an *agl24* mutant.

The effect of *AGL22* (also known as *SVP*) and *AGL24* loss-of-function mutations was assessed. An *agl24* T-DNA insertional mutant (designated W24.2) and an *agl22* mutant (designated *svp-E*) were obtained and the time to flowering of the mutant plants was measured and compared to wildtype Columbia Arabidopsis plants. On average, the *agl24* mutant produced almost twice as many leaves before flowering than wildtype plants. In addition, the *agl22* mutant produced only half the number of leaves as wildtype before flowering. Results of the experiment, shown in number of leaves prior to flowering, is provided below.

	Rosette	Cauline	Total	N
Columbia	11 +/- 0.9	2.9 +/- 0.5	14 +/- 1.1	26
<i>svp-E</i>	6 +/- 0.6	2.8 +/- 0.4	9 +/- 0.6	25
W24.2	19 +/- 1.5	3.1 +/- 0.5	22 +/- 1.7	26

Thus, the time to flowering and the amount of vegetative growth of the *agl24* mutant was increased compared to wild type plants and the time to flowering and the amount of vegetative growth of the *agl22* mutant was decreased compared to wild type plants.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, databases, Genbank sequences, patents, and patent applications cited herein are hereby incorporated by reference.

SEQUENCE LISTING

```
#          SEQUENCE LISTING          - - - - (1) GENERAL INFORMATION:
- -      (iii) NUMBER OF SEQUENCES: 26
- -      (2) INFORMATION FOR SEQ ID NO:1:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 1057 base - #pairs
- -          (B) TYPE: nucleic acid
- -          (C) STRANDEDNESS: double
- -          (D) TOPOLOGY: linear
- -          (ix) FEATURE:
- -              (A) NAME/KEY: CDS
- -              (B) LOCATION: 124..893
- -              (ix) FEATURE:
- -                  (A) NAME/KEY: misc.sub.-- - #feature
- -                  (B) LOCATION: 1..1057
- -                  (D) OTHER INFORMATION: - #/note= "product = Arabidopsis
- -                      thaliana - #AP1."
- -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
- -      CTTTCCAATT GGTTCATACC AAAGTCTGAG CTCTTCTTTA TATCTCTCTT GT -
#AGTTTCTT      60
- -      ATTGGGGGTC TTGTTTTGT TTGGTTCTTT TAGAGTAAGA AGTTTCTTAA AA -
#AAGGATCA      120
- -      AAA ATG GGA AGG GGT AGG GTT CAA TTG AAG AG - #G ATA GAG AAC AAG
ATC      168
- -      Met Gly Arg Gly Arg Val Gln Leu - #Lys Arg Ile Glu Asn Lys Ile
- -          1 - # 5 - # 10 - # 15
- -      AAT AGA CAA GTG ACA TTC TCG AAA AGA AGA GC - #T GGT CTT TTG AAG AAA
- -          216
- -      Asn Arg Gln Val Thr Phe Ser Lys Arg Arg Al - #a Gly Leu Leu Lys Lys
- -          20 - # 25 - # 30
- -      GCT CAT GAG ATC TCT GTT CTC TGT GAT GCT GA - #A GTT GCT CTT GTT GTC
- -          264
- -      Ala His Glu Ile Ser Val Leu Cys Asp Ala Gl - #u Val Ala Leu Val Val
- -          35 - # 40 - # 45
- -      TTC TCC CAT AAG GGG AAA CTC TTC GAA TAC TC - #C ACT GAT TCT TGT ATG
- -          312
- -      Phe Ser His Lys Gly Lys Leu Phe Glu Tyr Se - #r Thr Asp Ser Cys Met
- -          50 - # 55 - # 60
- -      GAG AAG ATA CTT GAA CGC TAT GAG AGG TAC TC - #T TAC GCC GAA AGA CAG
- -          360
- -      Glu Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Se - #r Tyr Ala Glu Arg Gln
- -          65 - # 70 - # 75
- -      CTT ATT GCA CCT GAG TCC GAC GTC AAT ACA AA - #C TGG TCG ATG GAG TAT
- -          408
- -      Leu Ile Ala Pro Glu Ser Asp Val Asn Thr As - #n Trp Ser Met Glu Tyr
- -          80 - # 85 - # 90 - # 95
- -      AAC AGG CTT AAG GCT AAG ATT GAG CTT TTG GA - #G AGA AAC CAG AGG CAT
- -          456
- -      Asn Arg Leu Lys Ala Lys Ile Glu Leu Leu Gl - #u Arg Asn Gln Arg His
- -          100 - # 105 - # 110
- -      TAT CTT GGG GAA GAC TTG CAA GCA ATG AGC CC - #T AAA GAG CTT CAG AAT
- -          504
- -      Tyr Leu Gly Glu Asp Leu Gln Ala Met Ser Pr - #o Lys Glu Leu Gln Asn
- -          115 - # 120 - # 125
```

- - CTG GAG CAG CAG CTT GAC ACT GCT CTT AAG CA - #C ATC CGC ACT AGA AAA
 552
 Leu Glu Gln Gln Leu Asp Thr Ala Leu Lys Hi - #s Ile Arg Thr Arg Lys
 130 - # 135 - # 140
 - - AAC CAA CTT ATG TAC GAG TCC ATC AAT GAG CT - #C CAA AAA AAG GAG AAG
 600
 Asn Gln Leu Met Tyr Glu Ser Ile Asn Glu Le - #u Gln Lys Lys Glu Lys
 145 - # 150 - # 155
 - - GCC ATA CAG GAG CAA AAC AGC ATG CTT TCT AA - #A CAG ATC AAG GAG AGG
 648
 Ala Ile Gln Glu Gln Asn Ser Met Leu Ser Ly - #s Gln Ile Lys Glu Arg
 160 1 - #65 1 - #70 1 -
 #75
 - - GAA AAA ATT CTT AGG GCT CAA CAG GAG CAG TG - #G GAT CAG CAG AAC
 CAA 696
 Glu Lys Ile Leu Arg Ala Gln Gln Glu Gln Tr - #p Asp Gln Gln Asn Gln
 180 - # 185 - # 190
 - - GGC CAC AAT ATG CCT CCC CCT CTG CCA CCG CA - #G CAG CAC CAA ATC CAG
 744
 Gly His Asn Met Pro Pro Pro Leu Pro Pro Gl - #n Gln His Gln Ile Gln
 195 - # 200 - # 205
 - - CAT CCT TAC ATG CTC TCT CAT CAG CCA TCT CC - #T TTT CTC AAC ATG GGT
 792
 His Pro Tyr Met Leu Ser His Gln Pro Ser Pr - #o Phe Leu Asn Met Gly
 210 - # 215 - # 220
 - - GGT CTG TAT CAA GAA GAT GAT CCA ATG GCA AT - #G AGG AGG AAT GAT CTC
 840
 Gly Leu Tyr Gln Glu Asp Asp Pro Met Ala Me - #t Arg Arg Asn Asp Leu
 225 - # 230 - # 235
 - - GAA CTG ACT CTT GAA CCC GTT TAC AAC TGC AA - #C CTT GGC TGC TTC GCC
 888
 Glu Leu Thr Leu Glu Pro Val Tyr Asn Cys As - #n Leu Gly Cys Phe Ala
 240 2 - #45 2 - #50 2 -
 #55
 - - GCA TG AAGCATTTCC ATATATATAT TTGTAATCGT CAACAATAAA AAC - #AGTTTGC
 943 Ala
 - - CACATACATA TAAATAGTGG CTAGGCTCTT TTCATCCAAT TAATATATTT TG -
 #GCAAATGT 1003
 - - TCGATGTTCT TATATCATCA TATATAAATT AGCAGGCTCC TTTCTTTTTT TG - #TA
 1057 - - - (2) INFORMATION FOR SEQ ID NO:2:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 256 amino - #acids (B) TYPE: amino acid
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
 - - Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Th - #r Asp Ser Cys Met Glu

197
 Cys Asp Ala Glu Val Ala Leu Val Val Phe Se - #r His Lys Gly Lys Leu
 40 - # 45 - # 50
 - - TTT GAA TAC TCC ACT GAT TCT TGT ATG GAG AA - #G ATA CTT GAA CGC TAT
 245
 Phe Glu Tyr Ser Thr Asp Ser Cys Met Glu Ly - #s Ile Leu Glu Arg Tyr
 55 - # 60 - # 65 - # 70
 - - GAG AGA TAC TCT TAC GCC GAG AGA CAG CTT AT - #A GCA CCT GAG TCC GAC
 293
 Glu Arg Tyr Ser Tyr Ala Glu Arg Gln Leu Il - #e Ala Pro Glu Ser Asp
 75 - # 80 - # 85
 - - TCC AAT ACG AAC TGG TCG ATG GAG TAT AAT AG - #G CTT AAG GCT AAG ATT
 341
 Ser Asn Thr Asn Trp Ser Met Glu Tyr Asn Ar - #g Leu Lys Ala Lys Ile
 90 - # 95 - # 100
 - - GAG CTT TTG GAG AGA AAC CAG AGG CAC TAT CT - #T GGG GAA GAC TTG CAA
 389
 Glu Leu Leu Glu Arg Asn Gln Arg His Tyr Le - #u Gly Glu Asp Leu Gln
 105 - # 110 - # 115
 - - GCA ATG AGC CCT AAG GAA CTC CAG AAT CTA GA - #G CAA CAG CTT GAT ACT
 437
 Ala Met Ser Pro Lys Glu Leu Gln Asn Leu Gl - #u Gln Gln Leu Asp Thr
 120 - # 125 - # 130
 - - GCT CTT AAG CAC ATC CGC TCT AGA AAA AAC CA - #A CTT ATG TAC GAC TCC
 485
 Ala Leu Lys His Ile Arg Ser Arg Lys Asn Gl - #n Leu Met Tyr Asp Ser
 135 1 - #40 1 - #45 1 -
 #50
 - - ATC AAT GAG CTC CAA AGA AAG GAG AAA GCC AT - #A CAG GAA CAA AAC
 AGC 533
 Ile Asn Glu Leu Gln Arg Lys Glu Lys Ala Il - #e Gln Glu Gln Asn Ser
 155 - # 160 - # 165
 - - ATG CTT TCC AAG CAG ATT AAG GAG AGG GAA AA - #C GTT CTT AGG GCG CAA
 581
 Met Leu Ser Lys Gln Ile Lys Glu Arg Glu As - #n Val Leu Arg Ala Gln
 170 - # 175 - # 180
 - - CAA GAG CAA TGG GAC GAG CAG AAC CAT GGC CA - #T AAT ATG CCT CCG CCT
 629
 Gln Glu Gln Trp Asp Glu Gln Asn His Gly Hi - #s Asn Met Pro Pro Pro
 185 - # 190 - # 195
 - - CCA CCC CCG CAG CAG CAT CAA ATC CAG CAT CC - #T TAC ATG CTC TCT CAT
 677
 Pro Pro Pro Gln Gln His Gln Ile Gln His Pr - #o Tyr Met Leu Ser His
 200 - # 205 - # 210
 - - CAG CCA TCT CCT TTT CTC AAC ATG GGG GGG CT - #G TAT CAA GAA GAA GAT
 725
 Gln Pro Ser Pro Phe Leu Asn Met Gly Gly Le - #u Tyr Gln Glu Glu Asp
 215 2 - #20 2 - #25 2 -
 #30
 - - CAA ATG GCA ATG AGG AGG AAC GAT CTC GAT CT - #G TCT CTT GAA CCC
 GGT 773

Gln Met Ala Met Arg Arg Asn Asp Leu Asp Le - #u Ser Leu Glu Pro Gly
235 - # 240 - # 245
- - TAT AAC TGC AAT CTC GGC TGC - # - #
794 Tyr Asn Cys Asn Leu Gly Cys 250
- - - (2) INFORMATION FOR SEQ ID NO:4:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 253 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
- - Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Met Lys Lys Ala
20 - # 25 - # 30
- - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
35 - # 40 - # 45
- - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Th - #r Asp Ser Cys Met Glu
50 - # 55 - # 60
- - Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
65 - # 70 - # 75 - # 80
- - Ile Ala Pro Glu Ser Asp Ser Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
85 - # 90 - # 95
- - Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
100 - # 105 - # 110
- - Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
115 - # 120 - # 125
- - Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Ser Arg Lys Asn
130 - # 135 - # 140
- - Gln Leu Met Tyr Asp Ser Ile Asn Glu Leu Gl - #n Arg Lys Glu Lys Ala
145 1 - #50 1 - #55 1 -
#60
- - Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg
Glu
165 - # 170 - # 175
- - Asn Val Leu Arg Ala Gln Gln Glu Gln Trp As - #p Glu Gln Asn His Gly
180 - # 185 - # 190
- - His Asn Met Pro Pro Pro Pro Pro Gln Gl - #n His Gln Ile Gln His
195 - # 200 - # 205
- - Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
210 - # 215 - # 220
- - Leu Tyr Gln Glu Glu Asp Gln Met Ala Met Ar - #g Arg Asn Asp Leu Asp
225 2 - #30 2 - #35 2 -
#40 - - Leu Ser Leu Glu Pro Gly Tyr Asn Cys Asn Le - #u Gly Cys
245 - # 250
- - - (2) INFORMATION FOR SEQ ID NO:5:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 768 base - #pairs (C) STRANDEDNESS: double
(B) TYPE: nucleic acid (ii) MOLECULE TYPE: cDNA
(D) TOPOLOGY: linear - - (ix) FEATURE:
- - (A) NAME/KEY: CDS
(B) LOCATION: 1..766 - - (ix) FEATURE:

(A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..768
 (D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
 var. botr - #ytis AP1."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG GGA AGG GGT AGG GTT CAG TTG AAG AGG AT - #A GAA AAC AAG ATC AAT
 48
 Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 AGA CAA GTG ACA TTC TCG AAA AGA AGA GCT GG - #T CTT ATG AAG AAA GCT
 96
 Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Met Lys Lys Ala
 20 - # 25 - # 30
 CAT GAG ATC TCT GTT CTG TGT GAT GCT GAA GT - #T GCG CTT GTT GTC TTC
 144
 His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
 35 - # 40 - # 45
 TCC CAT AAG GGG AAA CTC TTT GAA TAC CCC AC - #T GAT TCT TGT ATG GAG
 192
 Ser His Lys Gly Lys Leu Phe Glu Tyr Pro Th - #r Asp Ser Cys Met Glu
 50 - # 55 - # 60
 GAG ATA CTT GAA CGC TAT GAG AGA TAC TCT TA - #C GCC GAG AGA CAG CTT
 240
 Glu Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
 65 - # 70 - # 75 - # 80
 ATA GCA CCT GAG TCC GAC TCC AAT ACG AAC TG - #G TCG ATG GAG TAT AAT
 288
 Ile Ala Pro Glu Ser Asp Ser Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
 85 - # 90 - # 95
 AGG CTT AAG GCT AAG ATT GAG CTT TTG GAG AG - #A AAC CAG AGG CAC TAT
 336
 Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
 100 - # 105 - # 110
 CTT GGG GAA GAC TTG CAA GCA ATG AGC CCT AA - #G GAA CTC CAG AAT CTA
 384
 Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
 115 - # 120 - # 125
 GAG CAA CAG CTT GAT ACT GCT CTT AAG CAC AT - #C CGC TCT AGA AAA AAC
 432
 Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Ser Arg Lys Asn
 130 - # 135 - # 140
 CAA CTT ATG TAC GAC TCC ATC AAT GAG CTC CA - #A AGA AAG GAG AAA GCC
 480
 Gln Leu Met Tyr Asp Ser Ile Asn Glu Leu Gl - #n Arg Lys Glu Lys Ala
 145 1 - #50 1 - #55 1 -
 #60
 ATA CAG GAA CAA AAC AGC ATG CTT TCC AAG CA - #G ATT AAG GAG AGG
 GAA 528
 Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg Glu
 165 - # 170 - # 175
 AAC GTT CTT AGG GCG CAA CAA GAG CAA TGG GA - #C GAG CAG AAC CAT GGC

576
Asn Val Leu Arg Ala Gln Gln Glu Gln Trp As - #p Glu Gln Asn His Gly
180 - # 185 - # 190
- - CAT AAT ATG CCT CCG CCT CCA CCC CCG CAG CA - #G CAT CAA ATC CAG CAT
624
His Asn Met Pro Pro Pro Pro Pro Pro Gln Gl - #n His Gln Ile Gln His
195 - # 200 - # 205
- - CCT TAC ATG CTC TCT CAT CAG CCA TCT CCT TT - #T CTC AAC ATG GGA GGG
672
Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
210 - # 215 - # 220
- - CTG TAT CAA GAA GAA GAT CAA ATG GCA ATG AG - #G AGG AAC GAT CTC GAT
720
Leu Tyr Gln Glu Glu Asp Gln Met Ala Met Ar - #g Arg Asn Asp Leu Asp
225 2 - #30 2 - #35 2 -
#40
- - CTG TCT CTT GAA CCC GTT TAC AAC TGC AAC CT - #T GGC CGT CGC TGC T
766
Leu Ser Leu Glu Pro Val Tyr Asn Cys Asn Le - #u Gly Arg Arg Cys
245 - # 250 - # 255
- - GA - # - # - #
768 - - - - (2) INFORMATION FOR SEQ ID NO:6:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 255 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
- - Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Met Lys Lys Ala
20 - # 25 - # 30
- - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
35 - # 40 - # 45
- - Ser His Lys Gly Lys Leu Phe Glu Tyr Pro Th - #r Asp Ser Cys Met Glu
50 - # 55 - # 60
- - Glu Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
65 - # 70 - # 75 - # 80
- - Ile Ala Pro Glu Ser Asp Ser Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
85 - # 90 - # 95
- - Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
100 - # 105 - # 110
- - Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
115 - # 120 - # 125
- - Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Ser Arg Lys Asn
130 - # 135 - # 140
- - Gln Leu Met Tyr Asp Ser Ile Asn Glu Leu Gl - #n Arg Lys Glu Lys Ala
145 1 - #50 1 - #55 1 -
#60
- - Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg
Glu 165 - # 170 - # 175
- - Asn Val Leu Arg Ala Gln Gln Glu Gln Trp As - #p Glu Gln Asn His Gly
180 - # 185 - # 190

- - His Asn Met Pro Pro Pro Pro Pro Gln Gl - #n His Gln Ile Gln His
 195 - # 200 - # 205
 - - Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
 210 - # 215 - # 220
 - - Leu Tyr Gln Glu Glu Asp Gln Met Ala Met Ar - #g Arg Asn Asp Leu Asp
 225 2 - #30 2 - #35 2 -
 #40
 - - Leu Ser Leu Glu Pro Val Tyr Asn Cys Asn Le - #u Gly Arg Arg Cys
 245 - # 250 - # 255
 - - - (2) INFORMATION FOR SEQ ID NO:7:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1345 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 149..968 - - (ix) FEATURE:
 (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..1345
 (D) OTHER INFORMATION: - #/note= "product = Zea mays AP1."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
 - - GCACGAGTCC TCCTCCTCCT CGCATCCAC CCCACCCAC CTTCTCCTTA AA -
 #GCTACCTG 60
 - - CCTACCCGGC GGTTCGCGC CGCAATCGAT CGACCGGAAG AGAAAGAGCA GC -
 #TAGCTAGC 120
 - - TAGCAGATCG GAGCACGGCA ACAAGGCG ATG GGG CGC GGC AAG - #GTA CAG CTG
 172
 - # Met Gly Arg - #Gly Lys Val Gln Leu
 - # 1 - # 5
 - - AAG CGG ATA GAG AAC AAG ATA AAC CGG CAG GT - #G ACC TTC TCC AAG CGC
 220
 Lys Arg Ile Glu Asn Lys Ile Asn Arg Gln Va - #l Thr Phe Ser Lys Arg
 10 - # 15 - # 20
 - - CGG AAC GGC CTG CTC AAG AAG GCG CAC GAG AT - #C TCC GTC CTC TGC GAT
 268
 Arg Asn Gly Leu Leu Lys Lys Ala His Glu Il - #e Ser Val Leu Cys Asp
 25 - # 30 - # 35 - # 40
 - - GCC GAG GTC GCC GTC ATC GTC TTC TCC CCC AA - #G GGC AAG CTC TAC GAG
 316
 Ala Glu Val Ala Val Ile Val Phe Ser Pro Ly - #s Gly Lys Leu Tyr Glu
 45 - # 50 - # 55
 - - TAC GCC ACC GAC TCC CGC ATG GAC AAA ATT CT - #T GAA CGC TAT GAG CGA
 364
 Tyr Ala Thr Asp Ser Arg Met Asp Lys Ile Le - #u Glu Arg Tyr Glu Arg
 60 - # 65 - # 70
 - - TAT TCC TAT GCT GAA AAG GCT CTT ATT TCA GC - #T GAA TCT GAA AGT GAG
 412
 Tyr Ser Tyr Ala Glu Lys Ala Leu Ile Ser Al - #a Glu Ser Glu Ser Glu
 75 - # 80 - # 85
 - - GGA AAT TGG TGC CAC GAA TAC AGG AAA CTG AA - #G GCC AAA ATT GAG ACC
 460
 Gly Asn Trp Cys His Glu Tyr Arg Lys Leu Ly - #s Ala Lys Ile Glu Thr

90 - # 95 - # 100
 - - ATA CAA AAA TGC CAC AAG CAC CTG ATG GGA GA - #G GAT CTA GAG TCT TTG
 508
 Ile Gln Lys Cys His Lys His Leu Met Gly Gl - #u Asp Leu Glu Ser Leu
 105 1 - #10 1 - #15 1 -
 #20
 - - AAT CCC AAA GAG CTC CAG CAA CTA GAG CAG CA - #G CTG GAT AGC TCA
 CTG 556
 Asn Pro Lys Glu Leu Gln Gln Leu Glu Gln Gl - #n Leu Asp Ser Ser Leu
 125 - # 130 - # 135
 - - AAG CAC ATC AGA TCA AGG AAG AGC CAC CTT AT - #G GCC GAG TCT ATT TCT
 604
 Lys His Ile Arg Ser Arg Lys Ser His Leu Me - #t Ala Glu Ser Ile Ser
 140 - # 145 - # 150
 - - GAG CTA CAG AAG AAG GAG AGG TCA CTG CAG GA - #G GAG AAC AAG GCT CTG
 652
 Glu Leu Gln Lys Lys Glu Arg Ser Leu Gln Gl - #u Glu Asn Lys Ala Leu
 155 - # 160 - # 165
 - - CAG AAG GAA CTT GCG GAG AGG CAG AAG GCC GT - #C GCG AGC CGG CAG CAG
 700
 Gln Lys Glu Leu Ala Glu Arg Gln Lys Ala Va - #l Ala Ser Arg Gln Gln
 170 - # 175 - # 180
 - - CAG CAA CAG CAG CAG GTG CAG TGG GAC CAG CA - #G ACA CAT GCC CAG GCC
 748
 Gln Gln Gln Gln Gln Val Gln Trp Asp Gln Gl - #n Thr His Ala Gln Ala
 185 1 - #90 1 - #95 2 -
 #00
 - - CAG ACA AGC TCA TCA TCG TCC TCC TTC ATG AT - #G AGG CAG GAT CAG
 CAG 796
 Gln Thr Ser Ser Ser Ser Ser Ser Phe Met Me - #t Arg Gln Asp Gln Gln
 205 - # 210 - # 215
 - - GGA CTG CCG CCT CCA CAC AAC ATC TGC TTC CC - #G CCG TTG ACA ATG GGA
 844
 Gly Leu Pro Pro Pro His Asn Ile Cys Phe Pr - #o Pro Leu Thr Met Gly
 220 - # 225 - # 230
 - - GAT AGA GGT GAA GAG CTG GCT GCG GCG GCG GC - #G GCG CAG CAG CAG CAG
 892
 Asp Arg Gly Glu Glu Leu Ala Ala Ala Ala Al - #a Ala Gln Gln Gln Gln
 235 - # 240 - # 245
 - - CCA CTG CCG GGG CAG GCG CAA CCG CAG CTC CG - #C ATC GCA GGT CTG CCA
 940
 Pro Leu Pro Gly Gln Ala Gln Pro Gln Leu Ar - #g Ile Ala Gly Leu Pro
 250 - # 255 - # 260
 - - CCA TGG ATG CTG AGC CAC CTC AAT GCA T AAGG - #AGAGGG TCGATGAACA
 988 Pro Trp Met Leu Ser His Leu Asn Ala
 265 2 - #70
 - - CATCGACCTC CTCTCTCTCT CTCTCTCGTC ATGGATCATG ACGTACGCGT AC -
 #CATATGGT 1048
 - - TGCTGTGCCT GCCCCCATCG ATCGCGAGCA ATGGCAGCT CATGCAAGTG AT -
 #CATTGCTC 1108
 - - CCCGTTGGTT AAACCCTAGC CTATGTTTCAT GCGTCAGCA ACTAAGCTAA AC -

```

#TATTGTTA 1168
- - TGTTTGCAAG AAAGGGTAAA CCCGCTAGCT GTGTAATCTT GTCCAGCTAT CA -
#GTATGCTT 1228
- - GTTACTGCCC AGTTACCCCTT GAATCTAGCG GCGCTTTTGG TGAGAGGGTG CA -
#GTTTACTT 1288
- - TAAACATGGT TCGTGACTTG CTGTAAATAG TAGTATTAAT CGATTTGGGC AT - #CTAAA
1345 - - - - (2) INFORMATION FOR SEQ ID NO:8:
- - (i) SEQUENCE CHARACTERISTICS:
- - (A) LENGTH: 273 amino - #acids (B) TYPE: amino acid
- - (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
- - Met Gly Arg Gly Lys Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gl - #y Leu Leu Lys Lys Ala
15 20 - # 25 - # 30
- - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Val Ile Val Phe
35 - # 40 - # 45
- - Ser Pro Lys Gly Lys Leu Tyr Glu Tyr Ala Th - #r Asp Ser Arg Met Asp
50 - # 55 - # 60
- - Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Ala Leu
65 - # 70 - # 75 - # 80
- - Ile Ser Ala Glu Ser Glu Ser Glu Gly Asn Tr - #p Cys His Glu Tyr Arg
85 - # 90 - # 95
- - Lys Leu Lys Ala Lys Ile Glu Thr Ile Gln Ly - #s Cys His Lys His Leu
100 - # 105 - # 110
- - Met Gly Glu Asp Leu Glu Ser Leu Asn Pro Ly - #s Glu Leu Gln Gln Leu
115 - # 120 - # 125
- - Glu Gln Gln Leu Asp Ser Ser Leu Lys His Il - #e Arg Ser Arg Lys Ser
130 - # 135 - # 140
- - His Leu Met Ala Glu Ser Ile Ser Glu Leu Gl - #n Lys Lys Glu Arg Ser
145 1 - #50 1 - #55 1 -
#60
- - Leu Gln Glu Glu Asn Lys Ala Leu Gln Lys Gl - #u Leu Ala Glu Arg
165 - # 170 - # 175
- - Lys Ala Val Ala Ser Arg Gln Gln Gln Gln Gl - #n Gln Gln Val Gln Trp
180 - # 185 - # 190
- - Asp Gln Gln Thr His Ala Gln Ala Gln Thr Se - #r Ser Ser Ser Ser Ser
195 - # 200 - # 205
- - Phe Met Met Arg Gln Asp Gln Gln Gly Leu Pr - #o Pro Pro His Asn Ile
210 - # 215 - # 220
- - Cys Phe Pro Pro Leu Thr Met Gly Asp Arg Gl - #y Glu Glu Leu Ala Ala
225 2 - #30 2 - #35 2 -
#40
- - Ala Ala Ala Ala Gln Gln Gln Gln Pro Leu Pr - #o Gly Gln Ala Gln
245 - # 250 - # 255
- - Gln Leu Arg Ile Ala Gly Leu Pro Pro Trp Me - #t Leu Ser His Leu Asn
260 - # 265 - # 270 - - Ala
- - - - (2) INFORMATION FOR SEQ ID NO:9:
- - (i) SEQUENCE CHARACTERISTICS:
- - (A) LENGTH: 779 base - #pairs

```

```

(B) TYPE: nucleic acid (C) STRANDEDNESS: double
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
- - (ix) FEATURE: (A) NAME/KEY: CDS
(B) LOCATION: 10..775 - - (ix) FEATURE:
(A) NAME/KEY: unsure (B) LOCATION: 778..779
(D) OTHER INFORMATION: - #/note= "N = one or more
nucleotides. - #" - - (ix) FEATURE:
(A) NAME/KEY: misc.sub.-- - #feature
(B) LOCATION: 1..779
(D) OTHER INFORMATION: - #/note= "product = Arabidopsis
thaliana - #CAL."
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
- - TTAAGAGAA ATG GGA AGG GGT AGG GTT GAA TTG AAG - # AGG ATA GAG AAC
48
Met Gly Arg Gly Arg - #Val Glu Leu Lys Arg Ile Glu Asn
1 - # 5 - # 10
- - AAG ATC AAT AGA CAA GTG ACA TTC TCG AAA AG - #A AGA ACT GGT CTT TTG
96
Lys Ile Asn Arg Gln Val Thr Phe Ser Lys Ar - #g Arg Thr Gly Leu Leu
15 - # 20 - # 25
- - AAG AAA GCT CAG GAG ATC TCT GTT CTT TGT GA - #T GCC GAG GTT TCC CTT
144
Lys Lys Ala Gln Glu Ile Ser Val Leu Cys As - #p Ala Glu Val Ser Leu
30 - # 35 - # 40 - # 45
- - ATT GTC TTC TCC CAT AAG GGC AAA TTG TTC GA - #G TAC TCC TCT GAA TCT
192
Ile Val Phe Ser His Lys Gly Lys Leu Phe Gl - #u Tyr Ser Ser Glu Ser
50 - # 55 - # 60
- - TGC ATG GAG AAG GTA CTA GAA CGC TAC GAG AG - #G TAT TCT TAC GCC GAG
240
Cys Met Glu Lys Val Leu Glu Arg Tyr Glu Ar - #g Tyr Ser Tyr Ala Glu
65 - # 70 - # 75
- - AGA CAG CTG ATT GCA CCT GAC TCT CAC GTT AA - #T GCA CAG ACG AAC TGG
288
Arg Gln Leu Ile Ala Pro Asp Ser His Val As - #n Ala Gln Thr Asn Trp
80 - # 85 - # 90
- - TCA ATG GAG TAT AGC AGG CTT AAG GCC AAG AT - #T GAG CTT TTG GAG AGA
336
Ser Met Glu Tyr Ser Arg Leu Lys Ala Lys Il - #e Glu Leu Leu Glu Arg
95 - # 100 - # 105
- - AAC CAA AGG CAT TAT CTG CGA GAA GAG TTG GA - #A CCA ATG AGC CTC AAG
384
Asn Gln Arg His Tyr Leu Gly Glu Glu Leu Gl - #u Pro Met Ser Leu Lys
110 1 - #15 1 - #20 1 -
#25
- - GAT CTC CAA AAT CTG GAG CAG CAG CTT GAG AC - #T GCT CTT AAG CAC
ATT 432
Asp Leu Gln Asn Leu Glu Gln Gln Leu Glu Th - #r Ala Leu Lys His Ile
130 - # 135 - # 140
- - CGC TCC AGA AAA AAT CAA CTC ATG AAT GAG TC - #C CTC AAC CAC CTC CAA
480

```

Arg Ser Arg Lys Asn Gln Leu Met Asn Glu Se - #r Leu Asn His Leu Gln
 145 - # 150 - # 155
 - - AGA AAG GAG AAG GAG ATA CAG GAG GAA AAC AG - #C ATG CTT ACC AAA CAG
 528
 Arg Lys Glu Lys Glu Ile Gln Glu Glu Asn Se - #r Met Leu Thr Lys Gln
 160 - # 165 - # 170
 - - ATA AAG GAG AGG GAA AAC ATC CTA AAG ACA AA - #A CAA ACC CAA TGT GAG
 576
 Ile Lys Glu Arg Glu Asn Ile Leu Lys Thr Ly - #s Gln Thr Gln Cys Glu
 175 - # 180 - # 185
 - - CAG CTG AAC CGC AGC GTC GAC GAT GTA CCA CA - #G CCA CAA CCA TTT CAA
 624
 Gln Leu Asn Arg Ser Val Asp Asp Val Pro Gl - #n Pro Gln Pro Phe Gln
 190 1 - #95 2 - #00 2 -
 #05
 - - CAC CCC CAT CTT TAC ATG ATC GCT CAT CAG AC - #T TCT CCT TTC CTA
 AAT 672
 His Pro His Leu Tyr Met Ile Ala His Gln Th - #r Ser Pro Phe Leu Asn
 210 - # 215 - # 220
 - - ATG GGT GGT TTG TAC CAA GGA GAA GAC CAA AC - #G GCG ATG AGG AGG AAC
 720
 Met Gly Gly Leu Tyr Gln Gly Glu Asp Gln Th - #r Ala Met Arg Arg Asn
 225 - # 230 - # 235
 - - AAT CTG GAT CTG ACT CTT GAA CCC ATT TAC AA - #T TAC CTT GGC TGT TAC
 768
 Asn Leu Asp Leu Thr Leu Glu Pro Ile Tyr As - #n Tyr Leu Gly Cys Tyr
 240 - # 245 - # 250
 - - GCC GCT T GANN - # - # - # 779
 Ala Ala 255 - - - - (2) INFORMATION FOR SEQ ID NO:10:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 255 amino - #acids (B) TYPE: amino acid
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
 - - Met Gly Arg Gly Arg Val Glu Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Thr Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - Gln Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Val Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
 65 - # 70 - # 75 - # 80
 - - Ile Ala Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Met Glu
 85 - # 90 - # 95
 - - Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Le - #u Glu Arg Asn Gln Arg
 100 - # 105 - # 110
 - - His Tyr Leu Gly Glu Glu Leu Glu Pro Met Se - #r Leu Lys Asp Leu Gln
 115 - # 120 - # 125
 - - Asn Leu Glu Gln Gln Leu Glu Thr Ala Leu Ly - #s His Ile Arg Ser Arg
 130 - # 135 - # 140

- - Lys Asn Gln Leu Met Asn Glu Ser Leu Asn Hi - #s Leu Gln Arg Lys Glu
145 1 - #50 1 - #55 1 -
#60
- - Lys Glu Ile Gln Glu Glu Asn Ser Met Leu Th - #r Lys Gln Ile Lys
5 Glu 165 - # 170 - # 175
- - Arg Glu Asn Ile Leu Lys Thr Lys Gln Thr Gl - #n Cys Glu Gln Leu Asn
180 - # 185 - # 190
- - Arg Ser Val Asp Asp Val Pro Gln Pro Gln Pr - #o Phe Gln His Pro His
195 - # 200 - # 205
10 - - Leu Tyr Met Ile Ala His Gln Thr Ser Pro Ph - #e Leu Asn Met Gly Gly
210 - # 215 - # 220
- - Leu Tyr Gln Gly Glu Asp Gln Thr Ala Met Ar - #g Arg Asn Asn Leu Asp
225 2 - #30 2 - #35 2 -
#40
15 - - Leu Thr Leu Glu Pro Ile Tyr Asn Tyr Leu Gl - #y Cys Tyr Ala Ala
245 - # 250 - # 255
- - - (2) INFORMATION FOR SEQ ID NO:11:
- - (i) SEQUENCE CHARACTERISTICS:
20 (A) LENGTH: 756 base - #pairs
(B) TYPE: nucleic acid (C) STRANDEDNESS: double
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
- - (ix) FEATURE: (A) NAME/KEY: CDS
(B) LOCATION: 1..754 - - (ix) FEATURE:
25 (A) NAME/KEY: misc.sub.-- - #feature
(B) LOCATION: 1..756
(D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
CAL." - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
- - ATG GGA AGG GGT AGG GTT GAA ATG AAG AGG AT - #A GAG AAC AAG ATC AAC
48
30 Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - CGA CAA GTG ACG TTT TCG AAA AGA AGA GCT GG - #T CTT TTG AAG AAA GCC
96
35 Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
20 - # 25 - # 30
- - CAT GAG ATC TCG ATC CTT TGT GAT GCT GAG GT - #T TCC CTT ATT GTC TTC
144
His Glu Ile Ser Ile Leu Cys Asp Ala Glu Va - #l Ser Leu Ile Val Phe
35 - # 40 - # 45
40 - - TCC CAT AAG GGG AAA CTG TTC GAG TAC TCG TC - #T GAA TCT TGC ATG GAG
192
Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
50 - # 55 - # 60
- - AAG GTA CTA GAA CAC TAC GAG AGG TAC TCT TA - #C GCC GAG AAA CAG CTA
45 240
Lys Val Leu Glu His Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu
65 - # 70 - # 75 - # 80
- - AAA GTT CCA GAC TCT CAC GTC AAT GCA CAA AC - #G AAC TGG TCA GTG GAA
288
50 Lys Val Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Val Glu
85 - # 90 - # 95

- - TAT AGC AGG CTT AAG GCT AAG ATT GAG CTT TT - #G GAG AGA AAC CAA AGG
336

Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Le - #u Glu Arg Asn Gln Arg
100 - # 105 - # 110

- - CAT TAT CTG GGC GAA GAT TTA GAA TCA ATC AG - #C ATA AAG GAG CTA CAG
384

His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
115 - # 120 - # 125

- - AAT CTG GAG CAG CAG CTT GAC ACT TCT CTT AA - #A CAT ATT CGC TCG AGA
432

Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
130 - # 135 - # 140

- - AAA AAT CAA CTA ATG CAC GAG TCC CTC AAC CA - #C CTC CAA AGA AAG GAG
480

Lys Asn Gln Leu Met His Glu Ser Leu Asn Hi - #s Leu Gln Arg Lys Glu
145 1 - #50 1 - #55 1 -

#60

- - AAA GAA ATA CTG GAG GAA AAC AGC ATG CTT GC - #C AAA CAG ATA AGG
GAG 528

Lys Glu Ile Leu Glu Glu Asn Ser Met Leu Al - #a Lys Gln Ile Arg Glu
165 - # 170 - # 175

- - AGG GAG AGT ATC CTA AGG ACA CAT CAA AAC CA - #A TCA GAG CAG CAA AAC
576

Arg Glu Ser Ile Leu Arg Thr His Gln Asn Gl - #n Ser Glu Gln Gln Asn
180 - # 185 - # 190

- - CGC AGC CAC CAT GTA GCT CCT CAG CCG CAA CC - #G CAG TTA AAT CCT TAC
624

Arg Ser His His Val Ala Pro Gln Pro Gln Pr - #o Gln Leu Asn Pro Tyr
195 - # 200 - # 205

- - ATG GCA TCA TCT CCT TTC CTA AAT ATG GGT GG - #C ATG TAC CAA GGA GAA
672

Met Ala Ser Ser Pro Phe Leu Asn Met Gly Gl - #y Met Tyr Gln Gly Glu
210 - # 215 - # 220

- - TAT CCA ACG GCG GTG AGG AGG AAC CGT CTC GA - #T CTG ACT CTT GAA CCC
720

Tyr Pro Thr Ala Val Arg Arg Asn Arg Leu As - #p Leu Thr Leu Glu Pro
225 2 - #30 2 - #35 2 -

#40 - - ATT TAC AAC TGC AAC CTT GGT TAC TTT GCC GC - #A T GA

- # 756 Ile Tyr Asn Cys Asn Leu Gly Tyr Phe Ala Al - #a
245 - # 250

- - - (2) INFORMATION FOR SEQ ID NO:12:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 251 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear

- - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
- - Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15

- - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
20 - # 25 - # 30

- - His Glu Ile Ser Ile Leu Cys Asp Ala Glu Val - #1 Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Val Leu Glu His Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu
 65 - # 70 - # 75 - # 80
 - - Lys Val Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Val Glu
 85 - # 90 - # 95
 - - Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Le - #u Glu Arg Asn Gln Arg
 100 - # 105 - # 110
 - - His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
 115 - # 120 - # 125
 - - Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
 130 - # 135 - # 140
 - - Lys Asn Gln Leu Met His Glu Ser Leu Asn Hi - #s Leu Gln Arg Lys Glu
 145 1 - #50 1 - #55 1 -
 #60
 - - Lys Glu Ile Leu Glu Glu Asn Ser Met Leu Al - #a Lys Gln Ile Arg
 Glu 165 - # 170 - # 175
 - - Arg Glu Ser Ile Leu Arg Thr His Gln Asn Gl - #n Ser Glu Gln Gln Asn
 180 - # 185 - # 190
 - - Arg Ser His His Val Ala Pro Gln Pro Gln Pr - #o Gln Leu Asn Pro Tyr
 195 - # 200 - # 205
 - - Met Ala Ser Ser Pro Phe Leu Asn Met Gly Gl - #y Met Tyr Gln Gly Glu
 210 - # 215 - # 220
 - - Tyr Pro Thr Ala Val Arg Arg Asn Arg Leu As - #p Leu Thr Leu Glu Pro
 225 2 - #30 2 - #35 2 -
 #40 - - Ile Tyr Asn Cys Asn Leu Gly Tyr Phe Ala Al - #a
 245 - # 250
 - - - (2) INFORMATION FOR SEQ ID NO:13:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 756 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 1..451 - - (ix) FEATURE:
 (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..756
 (D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
 var. botr - #ytis CAL."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
 - - ATG GGA AGG GGT AGG GTT GAA ATG AAG AGG AT - #A GAG AAC AAG ATC AAC
 48
 Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - AGA CAA GTG ACG TTT TCG AAA AGA AGA GCT GG - #T CTT TTG AAG AAA GCC
 96
 Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - CAT GAG ATC TCG ATT CTT TGT GAT GCT GAG GT - #T TCC CTT ATT GTC TTC
 144

His Glu Ile Ser Ile Leu Cys Asp Ala Glu Va - #1 Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - TCC CAT AAG GGG AAA CTG TTC GAG TAC TCG TC - #T GAA TCT TGC ATG GAG
 192
 Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - AAG GTA CTA GAA CGC TAC GAG AGG TAC TCT TA - #C GCC GAG AAA CAG CTA
 240
 Lys Val Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu
 65 - # 70 - # 75 - # 80
 - - AAA GCT CCA GAC TCT CAC GTC AAT GCA CAA AC - #G AAC TGG TCA ATG GAA
 288
 Lys Ala Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Met Glu
 85 - # 90 - # 95
 - - TAT AGC AGG CTT AAG GCT AAG ATT GAG CTT TG - #G GAG AGG AAC CAA AGG
 336
 Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Tr - #p Glu Arg Asn Gln Arg
 100 - # 105 - # 110
 - - CAT TAT CTG GGA GAA GAT TTA GAA TCA ATC AG - #C ATA AAG GAG CTA CAG
 384
 His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
 115 - # 120 - # 125
 - - AAT CTG GAG CAG CAG CTT GAC ACT TCT CTT AA - #A CAT ATT CGC TCC AGA
 432
 Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
 130 - # 135 - # 140
 - - AAA AAT CAA CTA ATG CAC T AGTCCCTCAA CCACCTCCAA - #AGAAAGGAGA
 481 Lys Asn Gln Leu Met His 145 1 - #50
 - - AAGAAATACT GGAGGAAAAC AGCATGCTTG CCAAACAGAT AAAGGAGAGG GA -
 #GAGTATCC 541
 - - TAAGGACACA TCAAAACCAA TCAGAGCAGC AAAACCGCAG CCACCATGTA GC -
 #TCCTCAGC 601
 - - CGCAACCGCA GTTAAATCCT TACATGGCAT CATCTCCTTT CCTAAATATG GG -
 #TGGCATGT 661
 - - ACCAAGGAGA ATATCCAACG GCGGTGAGGA GGAACCGTCT CGATCTGACT CT -
 #TGAACCCA 721 - - TTTACAACCTG CAACCTTGGT TACTTTGCCG CATGA - #
 - # 756 - - - - (2) INFORMATION FOR SEQ ID NO:14:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 150 amino - #acids
 (B) TYPE: amino acid (D) TOPOLOGY: linear
 - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
 - - Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile
 Asn 1 5 - # 10 - # 15
 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - His Glu Ile Ser Ile Leu Cys Asp Ala Glu Va - #1 Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Val Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu

```

65          - # 70          - # 75          - # 80
- - Lys Ala Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Met Glu
          85 - #          90 - #          95
- - Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Tr - #p Glu Arg Asn Gln Arg
          100 - #          105 - #          110
- - His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
          115 - #          120 - #          125
- - Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
          130 - #          135 - #          140
10 - - Lys Asn Gln Leu Met His          145          1 - #50
- - - - (2) INFORMATION FOR SEQ ID NO:15:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 1500 base - #pairs
      (B) TYPE: nucleic acid          (C) STRANDEDNESS: double
15      (D) TOPOLOGY: linear          - - (ii) MOLECULE TYPE: cDNA
- - (ix) FEATURE:          (A) NAME/KEY: CDS
      (B) LOCATION: 72..1343          - - (ix) FEATURE:
      (A) NAME/KEY: misc.sub.-- - #feature
      (B) LOCATION: 1..1500
      (D) OTHER INFORMATION: - #/note= "product = Arabidopsis
          thaliana - #LEAFY (LFY)."
```

20

```

- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
- - AAAGCAATCT GCTCAAAGA GTAAAGAAAG AGAGAAAAG AGAGTGATAG AG -
#AGAGAGAG          60
25 - - AAAAATAGAT T ATG GAT CCT GAA GGT TTC ACG AGT - #GGC TTA TTC CGG
TGG          110
          Met Asp Pro - #Glu Gly Phe Thr Ser Gly Leu Phe Arg Trp
          1 - #          5 - #          10
- - AAC CCA ACG AGA GCA TTG GTT CAA GCA CCA CC - #T CCG GTT CCA CCT CCG
30          158
Asn Pro Thr Arg Ala Leu Val Gln Ala Pro Pr - #o Pro Val Pro Pro Pro
          15 - #          20 - #          25
- - CTG CAG CAA CAG CCG GTG ACA CCG CAG ACG GC - #T GCT TTT GGG ATG CGA
          206
35 Leu Gln Gln Gln Pro Val Thr Pro Gln Thr Al - #a Ala Phe Gly Met Arg
          30 - # 35          - # 40          - # 45
- - CTT GGT GGT TTA GAG GGA CTA TTC GGT CCA TA - #C GGT ATA CGT TTC TAC
          254
40 Leu Gly Gly Leu Glu Gly Leu Phe Gly Pro Ty - #r Gly Ile Arg Phe Tyr
          50 - #          55 - #          60
- - ACG GCG GCG AAG ATA GCG GAG TTA GGT TTT AC - #G GCG AGC ACG CTT GTG
          302
Thr Ala Ala Lys Ile Ala Glu Leu Gly Phe Th - #r Ala Ser Thr Leu Val
          65 - #          70 - #          75
45 - - GGT ATG AAG GAC GAG GAG CTT GAA GAG ATG AT - #G AAT AGT CTC TCT CAT
          350
Gly Met Lys Asp Glu Glu Leu Glu Glu Met Me - #t Asn Ser Leu Ser His
          80 - #          85 - #          90
- - ATC TTT CGT TGG GAG CTT CTT GTT GGT GAA CG - #G TAC GGT ATC AAA GCT
50          398
Ile Phe Arg Trp Glu Leu Leu Val Gly Glu Ar - #g Tyr Gly Ile Lys Ala
```

95 - # 100 - # 105
 - - GCC GTT AGA GCT GAA CGG AGA CGA TTG CAA GA - #A GAG GAG GAA GAG GAA
 446
 Ala Val Arg Ala Glu Arg Arg Arg Leu Gln Gl - #u Glu Glu Glu Glu Glu
 110 1 - #15 1 - #20 1 -
 #25
 - - TCT TCT AGA CGC CGT CAT TTG CTA CTC TCC GC - #C GCT GGT GAT TCC
 GGT 494
 Ser Ser Arg Arg Arg His Leu Leu Leu Ser Al - #a Ala Gly Asp Ser Gly
 130 - # 135 - # 140
 - - ACT CAT CAC GCT CTT GAT GCT CTC TCC CAA GA - #A GAT GAT TGG ACA GGG
 542
 Thr His His Ala Leu Asp Ala Leu Ser Gln Gl - #u Asp Asp Trp Thr Gly
 145 - # 150 - # 155
 - - TTA TCT GAG GAA CCG GTG CAG CAA CAA GAC CA - #G ACT GAT GCG GCG GGG
 590
 Leu Ser Glu Glu Pro Val Gln Gln Gln Asp Gl - #n Thr Asp Ala Ala Gly
 160 - # 165 - # 170
 - - AAT AAC GGC GGA GGA GGA AGT GGT TAC TGG GA - #C GCA GGT CAA GGA AAG
 638
 Asn Asn Gly Gly Gly Gly Ser Gly Tyr Trp As - #p Ala Gly Gln Gly Lys
 175 - # 180 - # 185
 - - ATG AAG AAG CAA CAG CAG CAG AGA CGG AGA AA - #G AAA CCA ATG CTG ACG
 686
 Met Lys Lys Gln Gln Gln Gln Arg Arg Arg Ly - #s Lys Pro Met Leu Thr
 190 1 - #95 2 - #00 2 -
 #05
 - - TCA GTG GAA ACC GAC GAA GAC GTC AAC GAA GG - #T GAG GAT GAC GAC
 GGG 734
 Ser Val Glu Thr Asp Glu Asp Val Asn Glu Gl - #y Glu Asp Asp Asp Gly
 210 - # 215 - # 220
 - - ATG GAT AAC GGC AAC GGA GGT AGT GGT TTG GG - #G ACA GAG AGA CAG AGG
 782
 Met Asp Asn Gly Asn Gly Gly Ser Gly Leu Gl - #y Thr Glu Arg Gln Arg
 225 - # 230 - # 235
 - - GAG CAT CCG TTT ATC GTA ACG GAG CCT GGG GA - #A GTG GCA CGT GGC AAA
 830
 Glu His Pro Phe Ile Val Thr Glu Pro Gly Gl - #u Val Ala Arg Gly Lys
 240 - # 245 - # 250
 - - AAG AAC GGC TTA GAT TAT CTG TTC CAC TTG TA - #C GAA CAA TGC CGT GAG
 878
 Lys Asn Gly Leu Asp Tyr Leu Phe His Leu Ty - #r Glu Gln Cys Arg Glu
 255 - # 260 - # 265
 - - TTC CTT CTT CAG GTC CAG ACA ATT GCT AAA GA - #C CGT GGC GAA AAA TGC
 926
 Phe Leu Leu Gln Val Gln Thr Ile Ala Lys As - #p Arg Gly Glu Lys Cys
 270 2 - #75 2 - #80 2 -
 #85
 - - CCC ACC AAG GTG ACG AAC CAA GTA TTC AGG TA - #C GCG AAG AAA TCA
 GGA 974
 Pro Thr Lys Val Thr Asn Gln Val Phe Arg Ty - #r Ala Lys Lys Ser Gly

290 - # 295 - # 300
 - - GCG AGT TAC ATA AAC AAG CCT AAA ATG CGA CA - #C TAC GTT CAC TGT TAC
 1022
 Ala Ser Tyr Ile Asn Lys Pro Lys Met Arg Hi - #s Tyr Val His Cys Tyr
 305 - # 310 - # 315
 - - GCT CTC CAC TGC CTA GAC GAA GAA GCT TCA AA - #T GCT CTC AGA AGA GCC
 1070
 Ala Leu His Cys Leu Asp Glu Glu Ala Ser As - #n Ala Leu Arg Arg Ala
 320 - # 325 - # 330
 - - TTT AAA GAA CGC GGT GAG AAC GTT GGC TCA TG - #G CGT CAG GCT TGT TAC
 1118
 Phe Lys Glu Arg Gly Glu Asn Val Gly Ser Tr - #p Arg Gln Ala Cys Tyr
 335 - # 340 - # 345
 - - AAG CCA CTT GTG AAC ATC GCT TGT CGT CAT GG - #C TGG GAT ATA GAC GCC
 1166
 Lys Pro Leu Val Asn Ile Ala Cys Arg His Gl - #y Trp Asp Ile Asp Ala
 350 3 - #55 3 - #60 3 -
 #65
 - - GTC TTT AAC GCT CAT CCT CGT CTC TCT ATT TG - #G TAT GTT CCA ACA
 AAG 1214
 Val Phe Asn Ala His Pro Arg Leu Ser Ile Tr - #p Tyr Val Pro Thr Lys
 370 - # 375 - # 380
 - - CTG CGT CAG CTT TGC CAT TTG GAG CGG AAC AA - #T GCG GTT GCT GCG GCT
 1262
 Leu Arg Gln Leu Cys His Leu Glu Arg Asn As - #n Ala Val Ala Ala Ala
 385 - # 390 - # 395
 - - GCG GCT TTA GTT GGC GGT ATT AGC TGT ACC GG - #A TCG TCG ACG TCT GGA
 1310
 Ala Ala Leu Val Gly Gly Ile Ser Cys Thr Gl - #y Ser Ser Thr Ser Gly
 400 - # 405 - # 410
 - - CGT GGT GGA TGC GGC GGC GAC GAC TTG CGT TT - #C TAGTTTGTT TGGGTAGTT
 G 1363 Arg Gly Gly Cys Gly Gly Asp Asp Leu Arg Ph - #e
 415 - # 420
 - - TGGTTTGT TT AGTCGTTATC CTAATTAAC ATTAGTCTTT AATTTAGTCT TC -
 #TTGGCTAA 1423
 - - TTTATTTTTC TTTTITGTC AAAACCTTIA ATTTGTTATG GCTAATTTGT TA -
 #TACACGCA 1483
 - - GTTTTCTTAA TGC GTTA - # - #
 - # 1500 - - - (2) INFORMATION FOR SEQ ID NO:16:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 424 amino - #acids (B) TYPE: amino acid
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
 - - Met Asp Pro Glu Gly Phe Thr Ser Gly Leu Ph - #e Arg Trp Asn Pro Thr
 1 5 - # 10 - # 15
 - - Arg Ala Leu Val Gln Ala Pro Pro Pro Val Pr - #o Pro Pro Leu Gln Gln
 20 - # 25 - # 30
 - - Gln Pro Val Thr Pro Gln Thr Ala Ala Phe Gl - #y Met Arg Leu Gly Gly
 35 - # 40 - # 45
 - - Leu Glu Gly Leu Phe Gly Pro Tyr Gly Ile Ar - #g Phe Tyr Thr Ala Ala

50 - # 55 - # 60
 - - Lys Ile Ala Glu Leu Gly Phe Thr Ala Ser Th - #r Leu Val Gly Met Lys
 65 - # 70 - # 75 - # 80
 - - Asp Glu Glu Leu Glu Glu Met Met Asn Ser Le - #u Ser His Ile Phe Arg
 85 - # 90 - # 95
 - - Trp Glu Leu Leu Val Gly Glu Arg Tyr Gly Il - #e Lys Ala Ala Val Arg
 100 - # 105 - # 110
 - - Ala Glu Arg Arg Arg Leu Gln Glu Glu Gl Gl - #u Glu Glu Ser Ser Arg
 115 - # 120 - # 125
 - - Arg Arg His Leu Leu Leu Ser Ala Ala Gly As - #p Ser Gly Thr His His
 130 - # 135 - # 140
 - - Ala Leu Asp Ala Leu Ser Gln Glu Asp Asp Tr - #p Thr Gly Leu Ser Glu
 145 1 - #50 1 - #55 1 -
 #60
 - - Glu Pro Val Gln Gln Gln Asp Gln Thr Asp Al - #a Ala Gly Asn Asn
 Gly 165 - # 170 - # 175
 - - Gly Gly Gly Ser Gly Tyr Trp Asp Ala Gly Gl - #n Gly Lys Met Lys Lys
 180 - # 185 - # 190
 - - Gln Gln Gln Gln Arg Arg Arg Lys Lys Pro Me - #t Leu Thr Ser Val Glu
 195 - # 200 - # 205
 - - Thr Asp Glu Asp Val Asn Glu Gly Glu Asp As - #p Asp Gly Met Asp Asn
 210 - # 215 - # 220
 - - Gly Asn Gly Gly Ser Gly Leu Gly Thr Glu Ar - #g Gln Arg Glu His Pro
 225 2 - #30 2 - #35 2 -
 #40
 - - Phe Ile Val Thr Glu Pro Gly Glu Val Ala Ar - #g Gly Lys Lys Asn
 Gly 245 - # 250 - # 255
 - - Leu Asp Tyr Leu Phe His Leu Tyr Glu Gln Cy - #s Arg Glu Phe Leu Leu
 260 - # 265 - # 270
 - - Gln Val Gln Thr Ile Ala Lys Asp Arg Gly Gl - #u Lys Cys Pro Thr Lys
 275 - # 280 - # 285
 - - Val Thr Asn Gln Val Phe Arg Tyr Ala Lys Ly - #s Ser Gly Ala Ser Tyr
 290 - # 295 - # 300
 - - Ile Asn Lys Pro Lys Met Arg His Tyr Val Hi - #s Cys Tyr Ala Leu His
 305 3 - #10 3 - #15 3 -
 #20
 - - Cys Leu Asp Glu Glu Ala Ser Asn Ala Leu Ar - #g Arg Ala Phe Lys
 Glu 325 - # 330 - # 335
 - - Arg Gly Glu Asn Val Gly Ser Trp Arg Gln Al - #a Cys Tyr Lys Pro Leu
 340 - # 345 - # 350
 - - Val Asn Ile Ala Cys Arg His Gly Trp Asp Il - #e Asp Ala Val Phe Asn
 355 - # 360 - # 365
 - - Ala His Pro Arg Leu Ser Ile Trp Tyr Val Pr - #o Thr Lys Leu Arg Gln
 370 - # 375 - # 380
 - - Leu Cys His Leu Glu Arg Asn Asn Ala Val Al - #a Ala Ala Ala Ala Leu
 385 3 - #90 3 - #95 4 -
 #00
 - - Val Gly Gly Ile Ser Cys Thr Gly Ser Ser Th - #r Ser Gly Arg Gly
 Gly 405 - # 410 - # 415
 - - Cys Gly Gly Asp Asp Leu Arg Phe 420
 - - - - (2) INFORMATION FOR SEQ ID NO:17:

- - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1656 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ix) FEATURE:
 (A) NAME/KEY: CDS (B) LOCATION: 1..1651
 - - (ix) FEATURE: (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..1656
 (D) OTHER INFORMATION: - #/note= "domain = ecdysone receptor
 ligand bi - #nding domain."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
 - - ATG CGG CCG GAA TGC GTC GTC CCG GAG AAC CA - #A TGT GCG ATG AAG CGG

48
 Met Arg Pro Glu Cys Val Val Pro Glu Asn Gl - #n Cys Ala Met Lys Arg
 1 5 - # 10 - # 15
 - - CGC GAA AAG AAG GCC CAG AAG GAG AAG GAC AA - #A ATG ACC ACT TCG CCG
 96
 Arg Glu Lys Lys Ala Gln Lys Glu Lys Asp Ly - #s Met Thr Thr Ser Pro
 20 - # 25 - # 30
 - - AGC TCT CAG CAT GGC GGC AAT GGC AGC TTG GC - #C TCT GGT GGC GGC CAA
 144
 Ser Ser Gln His Gly Gly Asn Gly Ser Leu Al - #a Ser Gly Gly Gly Gln
 35 - # 40 - # 45
 - - GAC TTT GTT AAG AAG GAG ATT CTT GAC CTT AT - #G ACA TGC GAG CCG CCC
 192
 Asp Phe Val Lys Lys Glu Ile Leu Asp Leu Me - #t Thr Cys Glu Pro Pro
 50 - # 55 - # 60
 - - CAG CAT GCC ACT ATT CCG CTA CTA CCT GAT GA - #A ATA TTG GCC AAG TGT
 240
 Gln His Ala Thr Ile Pro Leu Leu Pro Asp Gl - #u Ile Leu Ala Lys Cys
 65 - # 70 - # 75 - # 80
 - - CAA GCG CGC AAT ATA CCT TCC TTA ACG TAC AA - #T CAG TTG GCC GTT ATA
 288
 Gln Ala Arg Asn Ile Pro Ser Leu Thr Tyr As - #n Gln Leu Ala Val Ile
 85 - # 90 - # 95
 - - TAC AAG TTA ATT TGG TAC CAG GAT GGC TAT GA - #G CAG CCA TCT GAA GAG
 336
 Tyr Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Gl - #u Gln Pro Ser Glu Glu
 100 - # 105 - # 110
 - - GAT CTC AGG CGT ATA ATG AGT CAA CCC GAT GA - #G AAC GAG AGC CAA ACG
 384
 Asp Leu Arg Arg Ile Met Ser Gln Pro Asp Gl - #u Asn Glu Ser Gln Thr
 115 - # 120 - # 125
 - - GAC GTC AGC TTT CGG CAT ATA ACC GAG ATA AC - #C ATA CTC ACG GTC CAG
 432
 Asp Val Ser Phe Arg His Ile Thr Glu Ile Th - #r Ile Leu Thr Val Gln
 130 - # 135 - # 140
 - - TTG ATT GTT GAG TTT GCT AAA GGT CTA CCA GC - #G TTT ACA AAG ATA CCC
 480
 Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Al - #a Phe Thr Lys Ile Pro
 145 1 - #50 1 - #55 1 -
 #60

- - CAG GAG GAC CAG ATC ACG TTA CTA AAG GCC TG - #C TCG TCG GAG GTG
 ATG 528
 Gln Glu Asp Gln Ile Thr Leu Leu Lys Ala Cy - #s Ser Ser Glu Val Met
 165 - # 170 - # 175
 - - ATG CTG CGT ATG GCA CGA CGC TAT GAC CAC AG - #C TCG GAC TCA ATA TTC
 576
 Met Leu Arg Met Ala Arg Arg Tyr Asp His Se - #r Ser Asp Ser Ile Phe
 180 - # 185 - # 190
 - - TTC GCG AAT AAT AGA TCA TAT ACG CGG GAT TC - #T TAC AAA ATG GCC GGA
 624
 Phe Ala Asn Asn Arg Ser Tyr Thr Arg Asp Se - #r Tyr Lys Met Ala Gly
 195 - # 200 - # 205
 - - ATG GCT GAT AAC ATT GAA GAC CTG CTG CAT TT - #C TGC CGC CAA ATG TTC
 672
 Met Ala Asp Asn Ile Glu Asp Leu Leu His Ph - #e Cys Arg Gln Met Phe
 210 - # 215 - # 220
 - - TCG ATG AAG GTG GAC AAC GTC GAA TAC GCG CT - #T CTC ACT GCC ATT GTG
 720
 Ser Met Lys Val Asp Asn Val Glu Tyr Ala Le - #u Leu Thr Ala Ile Val
 225 2 - #30 2 - #35 2 -
 #40
 - - ATC TTC TCG GAC CGG CCG GGC CTG GAG AAG GC - #C CAA CTA GTC GAA
 GCG 768
 Ile Phe Ser Asp Arg Pro Gly Leu Glu Lys Al - #a Gln Leu Val Glu Ala
 245 - # 250 - # 255
 - - ATC CAG AGC TAC TAC ATC GAC ACG CTA CGC AT - #T TAT ATA CTC AAC CGC
 816
 Ile Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Il - #e Tyr Ile Leu Asn Arg
 260 - # 265 - # 270
 - - CAC TGC GGC GAC TCA ATG AGC CTC GTC TTC TA - #C GCA AAG CTG CTC TCG
 864
 His Cys Gly Asp Ser Met Ser Leu Val Phe Ty - #r Ala Lys Leu Leu Ser
 275 - # 280 - # 285
 - - ATC CTC ACC GAG CTG CGT ACG CTG GGC AAC CA - #G AAC GCC GAG ATG TGT
 912
 Ile Leu Thr Glu Leu Arg Thr Leu Gly Asn Gl - #n Asn Ala Glu Met Cys
 290 - # 295 - # 300
 - - TTC TCA CTA AAG CTC AAA AAC CGC AAA CTG CC - #C AAG TTC CTC GAG GAG
 960
 Phe Ser Leu Lys Leu Lys Asn Arg Lys Leu Pr - #o Lys Phe Leu Glu Glu
 305 3 - #10 3 - #15 3 -
 #20
 - - ATC TGG GAC GTT CAT GCC ATC CCG CCA TCG GT - #C CAG TCG CAC CTT
 CAG 1008
 Ile Trp Asp Val His Ala Ile Pro Pro Ser Va - #l Gln Ser His Leu Gln
 325 - # 330 - # 335
 - - ATT ACC CAG GAG GAG AAC GAG CGT CTC GAG CG - #G GCT GAG CGT ATG CGG
 1056
 Ile Thr Gln Glu Glu Asn Glu Arg Leu Glu Ar - #g Ala Glu Arg Met Arg
 340 - # 345 - # 350
 - - GCA TCG GTT GGG GGC GCC ATT ACC GCC GGC AT - #T GAT TGC GAC TCT GCC

1104
Ala Ser Val Gly Gly Ala Ile Thr Ala Gly Il - #e Asp Cys Asp Ser Ala
355 - # 360 - # 365
- - TCC ACT TCG GCG GCG GCA GCC GCG GCC CAG CA - #T CAG CCT CAG CCT CAG
1152
Ser Thr Ser Ala Ala Ala Ala Ala Gln Hi - #s Gln Pro Gln Pro Gln
370 - # 375 - # 380
- - CCC CAG CCC CAA CCC TCC TCC CTG ACC CAG AA - #C GAT TCC CAG CAC CAG
1200
Pro Gln Pro Gln Pro Ser Ser Leu Thr Gln As - #n Asp Ser Gln His Gln
385 3 - #90 3 - #95 4 -
#00
- - ACA CAG CCG CAG CTA CAA CCT CAG CTA CCA CC - #T CAG CTG CAA GGT
CAA 1248
Thr Gln Pro Gln Leu Gln Pro Gln Leu Pro Pr - #o Gln Leu Gln Gly Gln
405 - # 410 - # 415
- - CTG CAA CCC CAG CTC CAA CCA CAG CTT CAG AC - #G CAA CTC CAG CCA CAG
1296
Leu Gln Pro Gln Leu Gln Pro Gln Leu Gln Th - #r Gln Leu Gln Pro Gln
420 - # 425 - # 430
- - ATT CAA CCA CAG CCA CAG CTC CTT CCC GTC TC - #C GCT CCC GTG CCC GCC
1344
Ile Gln Pro Gln Pro Gln Leu Leu Pro Val Se - #r Ala Pro Val Pro Ala
435 - # 440 - # 445
- - TCC GTA ACC GCA CCT GGT TCC TTG TCC GCG GT - #C AGT ACG AGC AGC GAA
1392
Ser Val Thr Ala Pro Gly Ser Leu Ser Ala Va - #l Ser Thr Ser Ser Glu
450 - # 455 - # 460
- - TAC ATG GGC GGA AGT GCG GCC ATA GGA CCC AT - #C ACG CCG GCA ACC ACC
1440
Tyr Met Gly Gly Ser Ala Ala Ile Gly Pro Il - #e Thr Pro Ala Thr Thr
465 4 - #70 4 - #75 4 -
#80
- - AGC AGT ATC ACG GCT GCC GTT ACC GCT AGC TC - #C ACC ACA TCA GCG
GTA 1488
Ser Ser Ile Thr Ala Ala Val Thr Ala Ser Se - #r Thr Thr Ser Ala Val
485 - # 490 - # 495
- - CCG ATG GGC AAC GGA GTT GGA GTC GGT GTT GG - #G GTG GGC GGC AAC GTC
1536
Pro Met Gly Asn Gly Val Gly Val Gly Val Gl - #y Val Gly Gly Asn Val
500 - # 505 - # 510
- - AGC ATG TAT GCG AAC GCC CAG ACG GCG ATG GC - #C TTG ATG GGT GTA GCC
1584
Ser Met Tyr Ala Asn Ala Gln Thr Ala Met Al - #a Leu Met Gly Val Ala
515 - # 520 - # 525
- - CTG CAT TCG CAC CAA GAG CAG CTT ATC GGG GG - #A GTG GCG GTT AAG TCG
1632
Leu His Ser His Gln Gln Leu Ile Gly Gl - #y Val Ala Val Lys Ser
530 - # 535 - # 540
- - GAG CAC TCG ACG ACT GCA T AGCAG - # - #
1656 Glu His Ser Thr Thr Ala 545 5 - #50

```

- - - - (2) INFORMATION FOR SEQ ID NO:18:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 550 amino - #acids          (B) TYPE: amino acid
- -          (D) TOPOLOGY: linear          - -      (ii) MOLECULE TYPE: protein
5  - -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
- - Met Arg Pro Glu Cys Val Val Pro Glu Asn Gl - #n Cys Ala Met Lys Arg
1          5 - #          10 - #          15
- - Arg Glu Lys Lys Ala Gln Lys Glu Lys Asp Ly - #s Met Thr Thr Ser Pro
20         - #          25         - #          30
10 - - Ser Ser Gln His Gly Gly Asn Gly Ser Leu Al - #a Ser Gly Gly Gly Gln
35         - #          40         - #          45

- - Asp Phe Val Lys Lys Glu Ile Leu Asp Leu Me - #t Thr Cys Glu Pro Pro
50         - #          55         - #          60
15 - - Gln His Ala Thr Ile Pro Leu Leu Pro Asp Gl - #u Ile Leu Ala Lys Cys
65         - # 70         - # 75         - # 80
- - Gln Ala Arg Asn Ile Pro Ser Leu Thr Tyr As - #n Gln Leu Ala Val Ile
85 - #          90 - #          95
- - Tyr Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Gl - #u Gln Pro Ser Glu Glu
100        - #          105        - #          110
- - Asp Leu Arg Arg Ile Met Ser Gln Pro Asp Gl - #u Asn Glu Ser Gln Thr
115        - #          120        - #          125
- - Asp Val Ser Phe Arg His Ile Thr Glu Ile Th - #r Ile Leu Thr Val Gln
130        - # 135        - # 140
25 - - Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Al - #a Phe Thr Lys Ile Pro
145        1 - #50          1 - #55          1 -
#60
- - Gln Glu Asp Gln Ile Thr Leu Leu Lys Ala Cy - #s Ser Ser Glu Val
Met          165 - #          170 - #          175
30 - - Met Leu Arg Met Ala Arg Arg Tyr Asp His Se - #r Ser Asp Ser Ile Phe
180        - #          185        - #          190
- - Phe Ala Asn Asn Arg Ser Tyr Thr Arg Asp Se - #r Tyr Lys Met Ala Gly
195        - #          200        - #          205
- - Met Ala Asp Asn Ile Glu Asp Leu Leu His Ph - #e Cys Arg Gln Met Phe
210        - #          215        - #          220
35 - - Ser Met Lys Val Asp Asn Val Glu Tyr Ala Le - #u Leu Thr Ala Ile Val
225        2 - #30          2 - #35          2 -
#40
- - Ile Phe Ser Asp Arg Pro Gly Leu Glu Lys Al - #a Gln Leu Val Glu
40 Ala          245 - #          250 - #          255
- - Ile Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Il - #e Tyr Ile Leu Asn Arg
260        - #          265        - #          270
- - His Cys Gly Asp Ser Met Ser Leu Val Phe Ty - #r Ala Lys Leu Leu Ser
275        - #          280        - #          285
45 - - Ile Leu Thr Glu Leu Arg Thr Leu Gly Asn Gl - #n Asn Ala Glu Met Cys
290        - #          295        - #          300
- - Phe Ser Leu Lys Leu Lys Asn Arg Lys Leu Pr - #o Lys Phe Leu Glu Glu
305        3 - #10          3 - #15          3 -
#20
50 - - Ile Trp Asp Val His Ala Ile Pro Pro Ser Va - #l Gln Ser His Leu
Gln          325 - #          330 - #          335

```

- - Ile Thr Gln Glu Glu Asn Glu Arg Leu Glu Ar - #g Ala Glu Arg Met Arg
 340 - # 345 - # 350
 - - Ala Ser Val Gly Gly Ala Ile Thr Ala Gly Il - #e Asp Cys Asp Ser Ala
 355 - # 360 - # 365
 - - Ser Thr Ser Ala Ala Ala Ala Ala Gln Hi - #s Gln Pro Gln Pro Gln
 370 - # 375 - # 380
 - - Pro Gln Pro Gln Pro Ser Ser Leu Thr Gln As - #n Asp Ser Gln His Gln
 385 3 - #90 3 - #95 4 -
 #00
 - - Thr Gln Pro Gln Leu Gln Pro Gln Leu Pro Pr - #o Gln Leu Gln Gly
 Gln 405 - # 410 - # 415
 - - Leu Gln Pro Gln Leu Gln Pro Gln Leu Gln Th - #r Gln Leu Gln Pro Gln
 420 - # 425 - # 430
 - - Ile Gln Pro Gln Pro Gln Leu Leu Pro Val Se - #r Ala Pro Val Pro Ala
 435 - # 440 - # 445
 - - Ser Val Thr Ala Pro Gly Ser Leu Ser Ala Va - #l Ser Thr Ser Ser Glu
 450 - # 455 - # 460
 - - Tyr Met Gly Gly Ser Ala Ala Ile Gly Pro Il - #e Thr Pro Ala Thr Thr
 465 4 - #70 4 - #75 4 -
 #80
 - - Ser Ser Ile Thr Ala Ala Val Thr Ala Ser Se - #r Thr Thr Ser Ala
 Val 485 - # 490 - # 495
 - - Pro Met Gly Asn Gly Val Gly Val Gly Val Gl - #y Val Gly Gly Asn Val
 500 - # 505 - # 510
 - - Ser Met Tyr Ala Asn Ala Gln Thr Ala Met Al - #a Leu Met Gly Val Ala
 515 - # 520 - # 525
 - - Leu His Ser His Gln Glu Gln Leu Ile Gly Gl - #y Val Ala Val Lys Ser
 530 - # 535 - # 540
 - - Glu His Ser Thr Thr Ala 545 5 - #50
 - - - - (2) INFORMATION FOR SEQ ID NO:19:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 855 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ix) FEATURE:
 (A) NAME/KEY: CDS (B) LOCATION: 1..853
 - - (ix) FEATURE: (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..855
 (D) OTHER INFORMATION: - #/note= "domain = glucocorticoid
 receptor - #ligand binding domain."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
 - - ACA AAG AAA AAA ATC AAA GGG ATT CAG CAA GC - #C ACT GCA GGA GTC TCA
 48
 Thr Lys Lys Lys Ile Lys Gly Ile Gln Gln Al - #a Thr Ala Gly Val Ser
 1 5 - # 10 - # 15
 - - CAA GAC ACT TCG GAA AAT CCT AAC AAA ACA AT - #A GTT CCT GCA GCA TTA
 96
 Gln Asp Thr Ser Glu Asn Pro Asn Lys Thr Il - #e Val Pro Ala Ala Leu
 20 - # 25 - # 30
 - - CCA CAG CTC ACC CCT ACC TTG GTG TCA CTG CT - #G GAG GTG ATT GAA CCC
 144
 Pro Gln Leu Thr Pro Thr Leu Val Ser Leu Le - #u Glu Val Ile Glu Pro

35 - # 40 - # 45
 - - GAG GTG TTG TAT GCA GGA TAT GAT AGC TCT GT - #T CCA GAT TCA GCA TGG
 192
 Glu Val Leu Tyr Ala Gly Tyr Asp Ser Ser Va - #l Pro Asp Ser Ala Trp
 50 - # 55 - # 60
 - - AGA ATT ATG ACC ACA CTC AAC ATG TTA GGT GG - #G CGT CAA GTG ATT GCA
 240
 Arg Ile Met Thr Thr Leu Asn Met Leu Gly Gl - #y Arg Gln Val Ile Ala
 65 - # 70 - # 75 - # 80
 - - GCA GTG AAA TGG GCA AAG GCG ATA CTA GGC TT - #G AGA AAC TTA CAC CTC
 288
 Ala Val Lys Trp Ala Lys Ala Ile Leu Gly Le - #u Arg Asn Leu His Leu
 85 - # 90 - # 95
 - - GAT GAC CAA ATG ACC CTG CTA CAG TAC TCA TG - #G ATG TTT CTC ATG GCA
 336
 Asp Asp Gln Met Thr Leu Leu Gln Tyr Ser Tr - #p Met Phe Leu Met Ala
 100 - # 105 - # 110
 - - TTT GCC TTG GGT TGG AGA TCA TAC AGA CAA TC - #A AGC GGA AAC CTG CTC
 384
 Phe Ala Leu Gly Trp Arg Ser Tyr Arg Gln Se - #r Ser Gly Asn Leu Leu
 115 - # 120 - # 125
 - - TGC TTT GCT CCT GAT CTG ATT ATT AAT GAG CA - #G AGA ATG TCT CTA CCC
 432
 Cys Phe Ala Pro Asp Leu Ile Ile Asn Glu Gl - #n Arg Met Ser Leu Pro
 130 - # 135 - # 140
 - - TGC ATG TAT GAC CAA TGT AAA CAC ATG CTG TT - #T GTC TCC TCT GAA TTA
 480
 Cys Met Tyr Asp Gln Cys Lys His Met Leu Ph - #e Val Ser Ser Glu Leu
 145 1 - #50 1 - #55 1 -
 #60
 - - CAA AGA TTG CAG GTA TCC TAT GAA GAG TAT CT - #C TGT ATG AAA ACC
 TTA 528
 Gln Arg Leu Gln Val Ser Tyr Glu Glu Tyr Le - #u Cys Met Lys Thr Leu
 165 - # 170 - # 175
 - - CTG CTT CTC TCC TCA GTT GCT AAG GAA GGT CT - #G AAG AGC CAA GAG TTA
 576
 Leu Leu Leu Ser Ser Val Ala Lys Glu Gly Le - #u Lys Ser Gln Glu Leu
 180 - # 185 - # 190
 - - TTT GAT GAG ATT CGA ATG ACT TAT ATC AAA GA - #G CTA GGA AAA GCC ATC
 624
 Phe Asp Glu Ile Arg Met Thr Tyr Ile Lys Gl - #u Leu Gly Lys Ala Ile
 195 - # 200 - # 205
 - - GTC AAA AGG GAA GGG AAC TCC AGT CAG AAC TG - #G CAA CGG TTT TAC CAA
 672
 Val Lys Arg Glu Gly Asn Ser Ser Gln Asn Tr - #p Gln Arg Phe Tyr Gln
 210 - # 215 - # 220
 - - CTG ACA AAG CTT CTG GAC TCC ATG CAT GAG GT - #G GTT GAG AAT CTC CTT
 720
 Leu Thr Lys Leu Leu Asp Ser Met His Glu Va - #l Val Glu Asn Leu Leu
 225 2 - #30 2 - #35 2 -
 #40

```

- - ACC TAC TGC TTC CAG ACA TTT TTG GAT AAG AC - #C ATG AGT ATT GAA
TTC      768
Thr Tyr Cys Phe Gln Thr Phe Leu Asp Lys Th - #r Met Ser Ile Glu Phe
      245 - #      250 - #      255
- - CCA GAG ATG TTA GCT GAA ATC ATC ACT AAT CA - #G ATA CCA AAA TAT TCA
      816
Pro Glu Met Leu Ala Glu Ile Ile Thr Asn Gl - #n Ile Pro Lys Tyr Ser
      260 - #      265 - #      270
- - AAT GGA AAT ATC AAA AAG CTT CTG TTT CAT CA - #A AAA T GA
- #      855      Asn Gly Asn Ile Lys Lys Leu Leu Phe His Gl - #n Lys
      275 - #      280
- - - - (2) INFORMATION FOR SEQ ID NO:20:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 284 amino - #acids (B) TYPE: amino acid
      (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
- - Thr Lys Lys Lys Ile Lys Gly Ile Gln Gln Al - #a Thr Ala Gly Val Ser
1      5 - #      10 - #      15
- - Gln Asp Thr Ser Glu Asn Pro Asn Lys Thr Il - #e Val Pro Ala Ala Leu
20      20 - #      25 - #      30
- - Pro Gln Leu Thr Pro Thr Leu Val Ser Leu Le - #u Glu Val Ile Glu Pro
35      35 - #      40 - #      45
- - Glu Val Leu Tyr Ala Gly Tyr Asp Ser Ser Va - #l Pro Asp Ser Ala Trp
50      50 - #      55 - #      60
- - Arg Ile Met Thr Thr Leu Asn Met Leu Gly Gl - #y Arg Gln Val Ile Ala
25      65 - # 70 - # 75 - # 80
- - Ala Val Lys Trp Ala Lys Ala Ile Leu Gly Le - #u Arg Asn Leu His Leu
85 - #      90 - #      95
- - Asp Asp Gln Met Thr Leu Leu Gln Tyr Ser Tr - #p Met Phe Leu Met Ala
100 - #      105 - #      110
- - Phe Ala Leu Gly Trp Arg Ser Tyr Arg Gln Se - #r Ser Gly Asn Leu Leu
115 - #      120 - #      125
- - Cys Phe Ala Pro Asp Leu Ile Ile Asn Glu Gl - #n Arg Met Ser Leu Pro
130 - #      135 - #      140
35 - - Cys Met Tyr Asp Gln Cys Lys His Met Leu Ph - #e Val Ser Ser Glu Leu
145      1 - #50      1 - #55      1 -
#60
- - Gln Arg Leu Gln Val Ser Tyr Glu Glu Tyr Le - #u Cys Met Lys Thr
Leu      165 - #      170 - #      175
40 - - Leu Leu Leu Ser Ser Val Ala Lys Glu Gly Le - #u Lys Ser Gln Glu Leu
180 - #      185 - #      190
- - Phe Asp Glu Ile Arg Met Thr Tyr Ile Lys Gl - #u Leu Gly Lys Ala Ile
195 - #      200 - #      205
- - Val Lys Arg Glu Gly Asn Ser Ser Gln Asn Tr - #p Gln Arg Phe Tyr Gln
210 - #      215 - #      220
45 - - Leu Thr Lys Leu Leu Asp Ser Met His Glu Va - #l Val Glu Asn Leu Leu
225      2 - #30      2 - #35      2 -
#40
- - Thr Tyr Cys Phe Gln Thr Phe Leu Asp Lys Th - #r Met Ser Ile Glu
Phe      245 - #      250 - #      255
- - Pro Glu Met Leu Ala Glu Ile Ile Thr Asn Gl - #n Ile Pro Lys Tyr Ser

```

```

260      - #      265      - #      270
- - Asn Gly Asn Ile Lys Lys Leu Leu Phe His Gl - #n Lys
275      - #      280
- - - - (2) INFORMATION FOR SEQ ID NO:21:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 50 base - #pairs          (B) TYPE: nucleic acid
- -          (C) STRANDEDNESS: double              (D) TOPOLOGY: linear
- -      (ix) FEATURE:          (A) NAME/KEY: misc.sub.-- - #feature
- -          (B) LOCATION: 1..50
- -          (D) OTHER INFORMATION: - #/note= "element = copper inducible
- -              regulatory - #element (ACE1 binding site)."
```

5

```

- -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
- -      AGCTTAGCGA TCGCTCTTTT CCGCTGAACC GTTCCAGCAA AAAAGACTAG - #      50
- - - - (2) INFORMATION FOR SEQ ID NO:22:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 19 base - #pairs          (B) TYPE: nucleic acid
- -          (C) STRANDEDNESS: double              (D) TOPOLOGY: linear
- -      (ix) FEATURE:          (A) NAME/KEY: misc.sub.-- - #feature
- -          (B) LOCATION: 1..19
- -          (D) OTHER INFORMATION: - #/note= "element = tet operator."
```

10

```

- -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:
- -      ACTCTATCAG TGATAGAGT - # - #
- -      - # 19 - - - - (2) INFORMATION FOR SEQ ID NO:23:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 29 base - #pairs          (B) TYPE: nucleic acid
- -          (C) STRANDEDNESS: double              (D) TOPOLOGY: linear
- -      (ix) FEATURE:          (A) NAME/KEY: misc.sub.-- - #feature
- -          (B) LOCATION: 1..29
- -          (D) OTHER INFORMATION: - #/note= "element = ecdysone response
- -              element."
```

15

```

- -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
- -      GATCCGACAA GGGTTCAATG CACTTGTC A - # - #
- -      29 - - - - (2) INFORMATION FOR SEQ ID NO:24:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 371 base - #pairs
- -          (B) TYPE: nucleic acid
- -          (C) STRANDEDNESS: double              (D) TOPOLOGY: linear
- -      (ix) FEATURE:          (A) NAME/KEY: misc.sub.-- - #feature
- -          (B) LOCATION: 1..371
- -          (D) OTHER INFORMATION: - #/note= "element = heat shock
- -              inducible - #regulatory element (HSP81-1 promoter)."
```

20

```

- -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
- -      GTGGAGTCTC GAAACGAAAA GAACTTTCTG GAATTCGTTT GCTCACAAAG CT -
- -      #AAAAACGG      60
- -      - - TGTATTTCAT CGAAATACGG CGTCGTTTTC AAAGAACAAT CCAGAAATCA CT -
- -      #GGTTTTTC      120
- -      - - TTTATTTC AAAGAAGAGAC TAGAACTTTA TTTCTCCTCT ATAAAATCAC TT -
- -      #TGTTTTTC      180
- -      - - CCTCTCTTCT TCATAAATCA ACAAACAAT CACAAATCTC TCGAAACGCT CT -
- -      #CGAAGTTC      240
```

25

30

35

40

45

50


```

- - CAAATTTTCT CTTAGCATTC TCTTTCGTTT CTCGTTTGCG TTGAATCAAA GT -
#TCGTTGCG      300
- - ATGGCGGATG TTCAGATGGC TGATGCAGAG ACTTTTGCTT TCCAAGCTGA GA -
#TTAACCAG      360
5  - - CTTCTTAGCT T          - #          - #          - #          371
- - - - (2) INFORMATION FOR SEQ ID NO:25:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 29 base - #pairs          (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single              (D) TOPOLOGY: linear
10 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
- - GGATCCGGAT CAAAAATGGG AAGGGGTAG          - #          - #
      29          - - - - (2) INFORMATION FOR SEQ ID NO:26:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 30 base - #pairs          (B) TYPE: nucleic acid
15 (C) STRANDEDNESS: single              (D) TOPOLOGY: linear
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
- - GGATCCGCTG CGGCGAAGCA GCCAAGGTTG          - #          - #
      30

```

SEQ ID NO:27 and SEQ ID NO:28

Arabidopsis SEP1 cDNA and Arabidopsis SEP1 amino acid sequence

```

      20          40          60          80          100
ATGGGAAGAG GAAGAGTAGA GCTGAAGAGG ATAGAGAACA AAATCAACAG ACAAGTAACG TTTGCAAAGC GTAGGAACGG TTTGTGAAG AAAGCTTATG
TACCCTTCTC CTTCATCTCT CGACTTCTCC TATCTCTTGT TTTAGTTGTC TGTTCAATGC AAACGTTTCG CATCCTTGCC AAACAACCTC TTTCGAATAC
25 M G R G R V E L K R I E N K I N R Q V T F A K R R N G L L K K A Y>

      120          140          160          180          200
AATTGTCTGT TCTCTGTGAT GCTGAAGTTG CTCTCATCAT CTCTCCAAAC CGTGGAAAGC TCTATGAGTT TTGCAGCTCC TCAACATGCG TCAAGACACT
TTAACAGACA AGAGACACTA CGACTTCAAC GAGAGTAGTA GAAGAGGTTG GCACCTTTTC AGATACTCAA AACGTCGAGG AGTTTGATAG AGTTCTGTGA
30 E L S V L C D A E V A L I I F S N R G K L Y E F C S S S N M L K T L>

      220          240          260          280          300
TGATCGGTAC CAGAAATGCA GCTATGGATC CATTGAAGTC AACACAAAC CTGCCAAAGA ACTTGAGAAC AGCTACAGAG AATATCTGAA GCTTAAGGGT
ACTAGCCATG GTCTTTACGT CGATACCTAG GTAACCTCAG TTGTTGTTTG GACGTTTCT TGAACCTTTG TCGATGTCTC TTATAGACTT CGAATTCCTA
35 D R Y Q K C S Y G S I E V N N K P A K E L E N S Y R E Y L K L K G>

      320          340          360          380          400
AGATATGAGA ACCTTCAACG TCAACAGAGA AATCTTCTTG GGGAGGATTT AGGACCTTTG AATTCAAAGG AGTTAGAGCA GCTTGAGCGT CAACTGGACG
TCTATACTCT TGGAACTTGC AGTTGTCTCT TTAGAAGAAC CCTCCTTAA TCTCGGAAAC TTAAGTTTCC TCAATCTCGT CGAAGCTGCA GTTGACCTGC
40 R Y E N L Q R Q Q R N L L G E D L G P L N S K E L E Q L E R Q L D>

      420          440          460          480          500
GCTCTCTCAA GCAAGTTCCG TCCATCAAGA CACAGTACAT GCTTGACCAG CTCTCGGATC TTCAAAATAA AGAGCAAAATG TTGCTTGAAA CCAATAGAGC
CGAGAGAGTT CGTTCAAGCC AGGTAGTTCT GTGTATGTA CGAACTGGTC GAGAGCCTAG AAGTTTATT TCTCGTTTAC AACGAACCTT GGTATCTCTG
45 G S L K Q V R S I K T Q Y M L D Q L S D L Q N K E Q M L L E T N R A>

      520          540          560          580          600
TTTGCAATG AAGCTGGATG ATATGATTGG TGTGAGAAGT CATCATATGG GAGGATGGGA AGGCGGTGAA CAGAAATGTTA CCTACGCGCA TCATCAAGCT
AAACCGTTAC TTGACCTTAC TATACTAACC ACACCTCTCA GTAGTATACC CTCTACCTCT TCCGCCACTT GTCTTACAAT GGATGCGCGT AGTAGTTCGA
50 L A M K L D D M I G V R S H H M G G W E G G E Q N V T Y A H H Q A>

      620          640          660          680          700
CAGTCTCAGG GACTATACCA GCCTCTTGAA TGCAATCCAA CTCTGCAAAT GGGGTATGAT AATCCAGTAT GCTCTGAGCA AATCACTGCG ACAACACAAG
GTCAGAGTCC CTGATATGGT CGGAGAACTT ACCTTAGGTT GAGACGTTTA CCCCATACTA TTAGGTGATA CGAGACTCGT TTAGTGACGC TGTGTGTTTC
55 Q S Q G L Y Q P L E C N P T L Q M G Y D N P V C S E Q I T A T T Q>

      720          740
CTCAGGCGCA GCCGGGAAAC GGTACATTC CAGGATGGAT GCTCTGA
GAGTCCGCGT CGGCCCTTTG CCAATGTAAG GTCTACCTA CGAGACT
60 A Q A Q P G N G Y I P G W M L *>

```

SEQ ID NO:29 and SEQ ID NO:30

Arabidopsis SEP2 cDNA and Arabidopsis SEP2 amino acid sequence

```

      20          40          60          80          100
ATGGGAAGAG GAAGAGTAGA GCTCAAGAGG ATAGAGAACA AAATCAACAG ACAAGTGACG TTTGCTAAAC GTAGAAATGG TTTGCTGAAA AAAGCTTATG
5 TACCCCTTCTC CTTCTCATCT CGAGTTCTCC TATCTCTTGT TTTAGTTGTC TGTTCACGTC AAACGATTTC CATCTTTACC AAACGACTTT TTTGGAATAC
M G R G R V E L K R I E N K I N R Q V T F A K R R N G L L K K A Y>

      120          140          160          180          200
AGCTTTCTGT TCTCTGCGAT GCTGAAGTCT CTCTCATCGT CTTCTCCAAC CGTGGCAAGC TCTACGAGTT CTGCAGCACC TCCAACATGC TCAAGACACT
10 TCGAAAGACA AGAGACGCTA CGACTTCAGA GAGAGTAGCA GAAGAGGTTG GCACCGTTTC AGATGCTCAA GACGTCGTGG AGTTGTGTAC AGTTCTGTGA
E L S V L C D A E V S L I V F S N R G K L Y E F C S T S N M L K T L>

      220          240          260          280          300
GGAAAGGTAT CAGAAGTGTA GCTATGGCTC CATTGAAGTC AACACRAAC CTGCTAAAGA GCTTGAGAAC AGTACAGAG AGTACTTGAA GCTGAAAGGT
15 CTTTCCATA GTCTTCACAT CGATACCGAG GTAACCTCAG TTGTTGTTTG GACGATTCTT CGAAGTCTTG TCGATGTCTC TCATGAACCT CGACTTTCCA
E R Y Q K C S Y G S I E V N N K P A K E L E N S Y R E Y L K L K G>

      320          340          360          380          400
AGATATGAAA ATCTGCAACG TCAGCAGAGA AATCTTCTTG GAGAGGATCT TGGACCTCTG AATTCAAAGG AGCTAGAGCA GCTTGAGCGT CAACTAGACG
20 TCTATACITT TAGACGTTGC AGTCGTCTCT TTAGAAGAAC CTCTCCTAGA ACCTGGAGAC TTAAGTTTCC TCGATCTCGT CGAAGTCGCA GTTGATCTGC
R Y E N L Q R Q Q R N L L G E D L G P L N S K E L E Q L E R Q L D>

      420          440          460          480          500
GCTCTCTGAA GCAAGTTGCG TGCATCAAGA CACAGTATAT GCTTGACCAG CTCTCTGATC TTCAAGGTAA GGAGCATATC TTGCTTGATG CCAACAGAGC
25 CGAGAGACTT CGTTCAAGCG ACGTAGTTCT GTGTATATA CGAAGTGTCT GAGAGACTAG AAGTTCCATT CCTCGTATAG AACGAACTAC GGTGTGTCTG
G S L K Q V R C I K T Q Y M L D Q L S D L Q G K E H I L L D A N R A>

      520          540          560          580          600
TTTGTCAATG AAGCTGGAAG ATATGATCGG CGTGAGACAT CACCATATAG GAGGAGGATG GGAAGGTGGT GATCAACAGA ATATTGCCTA TGGACATCCT
30 AAACAGTTAC TTCGACCTTC TATACTAGCC GCACCTCTGA GTGTGTATATC CTCTCCTAC CCTCCACCA CTAGTTGTCT TATAACGGAT ACCTGTAGGA
L S M K L E D M I G V R H H H I G G G W E G G D Q Q N I A Y G H P>

      620          640          660          680          700
CAGGCTCATT CTCAGGACT ATACCAATCT CTTGAATGTG ATCCCACCTT GCAAAITGGA TATAGCCATC CAGTGTGCTC AGAGCAAATG GCTGTGACGG
35 GTCGAGTAA GAGTCCCTGA TATGTTTAGA GAACTTACAC TAGGGTGAAG CTTTAACTT ATATCGGTAG GTCACACGAG TCTCGTTTAC CGACACTGCC
Q A H S Q G L Y Q S L E C D P T L Q I G Y S H P V C S E Q M A V T>

      720          740
TGCAAGGTCA GTCCCAACAA GGAAACGGCT ACATCCCTGG CTGGATGCTG TGA
40 ACGTTCCAGT CAGGTTTGT CTTTGCCTGA TGTAGGAGCC GACCTACGAC ACT
V Q G Q S Q Q G N G Y I P G W M L *>
```

SEQ ID NO:31 and SEQ ID NO:32

Arabidopsis SEP3 cDNA and Arabidopsis SEP3 amino acid sequence

```

      20          40          60          80          100
ATGGGAAGAG GGAGAGTAGA ATTGAAGAGG ATAGAGAACA AGATCAATAG GCAAGTGACG TTTGCAAAGA GAAGGAATGG TCTTTTGAAG AAAGCATACG
45 TACCCTTCTC CCTCTCATCT TAACCTTCCC TATCTCTTGT TCTAGTTATC CGTTCACGTC AAACGTTTCT CTTCTTACC AGAAAACTTC TTTCTATGTC
M G R G R V E L K R I E N K I N R Q V T F A K R R N G L L K K A Y>

      120          140          160          180          200
AGCTTTTCACT TCTATGTGAT GCAGAAGTTG CTCTCATCAT CTTCTCAAAT AGAGGAAAGC TGTACGAGTT TTGAGTAGT TCGAGCATGC TTCGACACT
50 TCGAAAGTCA AGATACACTA CGTCTTCAAC GAGAGTAGTA GAAGAGTTTA TCTCTTTTCG ACATGCTCAA AACGTCATCA AGCTCGTACG AAGCCTGTGA
E L S V L C D A E V A L I I F S N R G K L Y E F C S S S S M L R T L>

      220          240          260          280          300
GGAGAGGTAC CAAAGTGTA ACTATGGAGC ACCAGAAGCC AATGTGCTTT CAAGAGAGGC CTTAGCAGTT GAACCTAGTA GCCAGCAGGA GTATCTCAAG
55 CCTCTCCATG GTTTTACAT TGATACCTCG TGGTCTTGGG TTACACGGAA GTTCTCTCCG GAATCGTCAA CTTGAATCAT CGGTGCTCCT CATAGAGTTC
E R Y Q K C N Y G A P E P N V P S R E A L A V E L S S Q Q E Y L K>

      320          340          360          380          400
CTTAAGGAGC GTTATGAGC CTTACAAAGA ACCCAAGGA ATCTGTTGGG AGAAGATCTT GGACCTCTAA GTACAAAGGA GCTTGAGTCA CTTGAGAGAC
60 GAATTCCTCG CAATACCTCG GAATGTTTCT TGGGTTTCT TAGACAACCC TCTTCTAGAA CCTGGAGATT CATGTTTCTT CGAAGTCACT GAACCTCTTG
L K E R Y D A L Q R T Q R N L L G E D L G P L S T K E L E S L E R>

      420          440          460          480          500
AGCTTGATTC TTCCTTGAAG CAGATCAGAG CTCTCAGGAC ACAGTTTATG CTTGACCAGC TCAACGATCT TCAGAGTAAG TTAGCTGATG GGTATCAGAT
65 TCGAACTAAG AAGGAACCTC GTCTAGTCTC GAGAGTCTCG TGTCAAATAC GAACTGGTCG AGTTGCTAGA AGTCTCAATC AATCGACTAC CCATAGTCTA
```

Q L D S S L K Q I R A L R T Q F M L D Q L N D L Q S K L A D G Y Q M>

520 540 560 580 600

GCCACTCCAG CTGAACCCTA ACCAAGAAGA GGTGATCAC TACGGTCGTC ATCATCATCA ACAACAACAA CACTCCCAAG CTTTCTTCCA GCCTTTGGAA

CGGTGAGGTC GACTTGGGAT TGGTTCTTCT CCAACTAGTG ATGCCAGCAG TAGTAGTAGT TGTTGTTGTT GTGAGGGTTC GAAAGAAGGT CGGAAACCTT

P L Q L N P N Q E E V D H Y G R H H H Q Q Q Q H S Q A F F Q P L E>

620 640 660 680 700

TGTGAACCCA TTCTTCAGAT CGGTATCAG GGGCAGCAAG ATGGAATGGG AGCAGGACCA AGTGTGAATA ATTACATGTT GGGTTGGTTA CCTTATGACA

ACACTTGGGT AAGAAGTCTA GCCCATAGTC CCCGTCGTTT TACCTTACCC TCGTCCTGGT TCACACTTAT TAATGTACAA CCCAACCAAT GGAATACTGT

C E P I L Q I G Y Q G Q Q D G M G A G P S V N N Y M L G W L P Y D>

CCAACCTCTAT TTGA

GGTTGAGATA AACT

T N S I *>

SEQ ID NO:33 and SEQ ID NO:34

Arabidopsis AGL20 cDNA and Arabidopsis AGL20 amino acid sequence

20 40 60 80 100

ATGTGTGAGGG GCAAAACTCA GATGAAGAGA ATAGAGAATG CAACAAGCAG ACAAGTGACT TTCTCCAAAA GAAGGAATGG TTTGTTGAAG AAAGCCTTTG

TACCACTCCC CGTTTTGAGT CTACTTCTCT TATCTCTTAC GTTGTTCGTC TGTTCACTGA AAGAGGTTTT CTTCTTACC AAACAACCTC TTTCGGAAAC

M V R G K T Q M K R I E N A T S R Q V T F S K R R N G L L K K A F>

120 140 160 180 200

AGCTCTCAGT GCTTTGTGAT GCTGAAGTTT CTCTTATCAT CTCTCTCTCT AAAGGCAAAAC TTTATGAATT CGCCAGCTCC AATATGCAAG ATACCATAGA

TCGAGAGTCA CGAAACACTA CGACTTCAAA GAGAATAGTA GAAGAGAGGA TTTCGGTTTG AAATACTTAA GCGGTCGAGG TTATACGTTT TATGTTATCT

E L S V L C D A E V S L I I F S P K G K L Y E F A S S N M Q D T I D>

220 240 260 280 300

TCGTTATCTG AGGCATACTA AGGATCGAGT CAGCACCAAA CGGTTTCTG AAGAAAATAT GCAGCATTGG AAATATGAAG CAGCAAAACAT GATGAAGAAA

AGCAATAGAC TCCGTATGAT TCCTAGCTCA GTCTGTGTTT GGCCAAAGAC TTCTTTTATA CGTCGTAAC TTTTACTTTC GTCTGTTGTA CTACTTCTTT

R Y L R H T K D R V S T K P V S E E N M Q H L K Y E A A N M M K K>

320 340 360 380 400

ATTGAACAAC TCGAAGCTTC TAAACGTAAA CTCTTGGGAG AAGGCATAGG AACATGCTCA ATCGAGGAGC TGCAACAGAT TGAGCAACAG CTTGAGAAAA

TAACITGTTG AGCTTCGAAG ATTTGCATTT GAGAACCCCT TTCCGTATCC TTGTACGAGT TAGCTCCTCG ACGTTGTCTA ACTCGTTGTC GAACITCTTT

I E Q L E A S K R K L L G E G I G T C S I E E L Q Q I E Q Q L E K>

420 440 460 480 500

GTGTCAAATG TATTGAGCA AGAAAGACTC AAGTGTITAA GGAACAAATT GAGCAGCTCA AGCAAAAGGA GAAAGCTCTA GCTGCAGAAA ACGAGAAGCT

CACAGTTTAC ATAAGCTGTT TCTTTCTGAG TTCACAAATT CCTTGTITAA CTCGTCGAGT TCGTTTCTCT CTTTCGAGAT CGACGCTCTT TGCTCTTCGA

S V K C I R A R K T Q V P K E Q I E Q L K Q K E K A L A A E N E K L>

520 540 560 580 600

CTCTGAAAGG TGGGGATCTC ATGAAAGCGA AGTTTGGTCA AATAAGAATC AAGAAAGTAC TGGAAGAGGT GATGAAGAGA GTAGCCCAAG TTCTGAAGTA

GAGACTTTTC ACCCCTAGAG TACTTTCGTT TCAAACCACT TTATCTTAGT TTCTTTCATG ACCTTCTCCA CTACTTCTCT CATCGGTTTC AAGACTTCAT

S E K W G S H E S E V W S N K N Q E S T G R G D E E S S P S S E V>

620 640

GAGACGCAAT TGTTCAATGG GTTACCTTGT TCTTCAAGAA AGTGA

CTCTGCGTTA ACAAGTAACC CAATGGAACA AGAAGTTCTT TCACT

E T Q L F I G L P C S S R K *>

SEQ ID NO:35 and SEQ ID NO:36

Arabidopsis AGL22 cDNA and Arabidopsis AGL22 amino acid sequence

20 40 60 80 100

ATGGCGAGAG AAAAGATTCA GATCAGGAAG ATGACACAAG CAACGGCGAG ACAAGTGACG TTTTCGAAAC GAAGAAGAGG GCTTTTCAAG AAAGCTGAAG

TACCGTCTCT TTTTCTAAGT CTAGTCTCTC TAGCTGTGTC GTTGCCGCTC TGTTCACTGC AAAAGCTTTG CTTCTCTCC CGAAAAGTTC TTTCGACTTC

M A R E K I Q I R K I D N A T A R Q V T F S K R R R G L F K K A E>

120 140 160 180 200

AACTCTCCGT TCTCTGCGAC GCCGATGTCG CTCTCATCAT CTCTCTTCC ACCGGAAAAC TGTTGAGATT CTGTAGCTCC AGCATGAAGG AAGTCTCTAG

TTGAGAGGCA AGAGACGCTG CGGCTACAGC GAGAGTAGTA GAAGAGAAAG TGCCCTTTTG ACAAGCTCAA GACATCGAGG TCGTACTTCC TTTCGATCT

E L S V L C D A D V A L I I F S S T G K L F E F C S S S M K E V L E>

220 240 260 280 300

GAGGCATAAC TTNCAGTCAA AGAACTTGGA GAAGCTTCAT CAGCCATCTC TTGAGTTACA GCTGGTTGAG AACAGTGATC ACGCCCGAAT GAGTAAAGAA
CTCCGTATTG AANGTCAGTT TCTTGAACCT CTTCGAAGTA GTCGGTAGAG AACTCAATGT CGACCAACTC TTGTCAGTAG TCGGGGCTTA CTCATTCTTT
R H N X Q S K N L E K L H Q P S L E L Q L V E N S D H A R M S K E>

320 340 360 380 400
ATTGCGGACA AGAGCCACCG ACTAAGGCAA ATGAGAGGAG AGGAACTTCA AGGACTTGAC ATTGAAGAGC TTCAGCAGCT AGAGAAGGCC CTTGAAACTG
TAACGCCTGT TCTCGTGGC TGATTCCGTT TACTCTCCTC TCCTTGAAGT TCCTGAACGT TAACCTCTCG AAGTCGTCGA TCTCTCCGG GAACTTTGAC
I A D K S H R L R Q M R G E E L Q G L D I E E L Q Q L E K A L E T>

420 440 460 480 500
GTTTGACGCG TGTGATTGAA ACAAAGAGTG ACAAGATTAT GAGTGAGATC AGCGAACTTC AGAAAAAGGG AATGCAATTG ATGGATGAGA ACAAGCGGTT
CAAACCTGCGC ACACTAACTT TGTTTCTCAC TGTTCTAATA CTCACTCTAG TCGCTTGAAG TCTTTTCCCT TTACGTTAAC TACCTACTCT TGTTCCGCAA
G L T R V I E T K S D K I M S E I S E L Q K K G M Q L M D E N K R L>

520 540 560 580 600
GAGGCAGCAA GTATGTGTCT TACCCTCTCT GTTGATAACA AATCCCTTTC TTTTGTCTAC CATTAAACGTA CACACTCCTA AATTAAATCC CCAGTTGTCT
CTCCGTCGTT CATACACAGA ATGGGAGAGA CAACTATTGT TTAGGAAAG AAAACAGATG GTAATTGCAT GTGTGAGGAT TTAAATTAGG GGTCAACAGA
R Q Q V C V L P S L L I T N P F L L S T I N V H T P K F N P Q L S>

620
ACAACACATA TGTTTGATCA TACTGTGAGA TAA
TGTTGTGTAT ACAAACTAGT ATGACACTCT ATT
T T H M F D H T V R *>

SEQ ID NO:37 and SEQ ID NO:38

Arabidopsis AGL24 cDNA and Arabidopsis AGL24 amino acid sequence

20 40 60 80 100
ATGGCGAGAG AGAAGATAAG GATAAGAAG ATTGATAACA TAACAGCGAG ACAAGTTACT TTCTCAAAGA GAAGAAGAGG AATCTTCAAG AAAGCCGATG
TACCCTCTCT TCTTCTATTC CTATTTCTTC TAACATTATTG ATTGTCGCTC TGTTCAATGA AAGAGTTTCT CTTCTTCTCC TTAGAAGTTC TTTCGGCTAC
M A R E K I R I K K I D N I T A R Q V T F S K R R R G I F K K A D>

120 140 160 180 200
AACTTTCAGT TCTTTGCGAT GCTGATGTTG CTCTCATCAT CTTCTCTGCC ACCGGAAAGC TCTTCGAGTT CTCCAGCTCA AGAATGAGAG ACATATTGGG
TTGAAAGTCA AGAAACGCTA CGACTACAAAC GAGAGTAGTA GAAGAGACGG TGCCCTTTCC AGAAGCTCAA GAGGTCGAGT TCTTACTCTC TGTATAACCC
E L S V L C D A D V A L I I F S A T G K L F E F S S S R M R D I L G>

220 240 260 280 300
AAGGTATAGT CTTTCATCAA GTAACATCAA CAAATTGATG GATCCACCTT CTACTCATCT CCGGCTTGAG AATTGTAACC TCTCCAGACT AAGTAAGGAA
TTCCATATCA GAAGTACGTT CATTGTGATT GTTTAACTAC CTAGGTGGAA GATGAGTAGA GGCCGAACCT TTAACATTGG AGAGGTCTGA TTCATTCCCT
R Y S L H A S N I N K L M D P P S T H L R L E N C N L S R L S K E>

320 340 360 380 400
GTGGAAGACA AAACCAAGCA GCTACGGAAC CTGAGAGGAG AGGATCTTGA TGGATTGAAC TTAGAAGAGT TGCAGCGGCT GGAGAAACTA CTTGAATCCG
CAGCTTCTGT TTTGGTTCTG CGATGCTCTT GACTCTCCTC TCCTAGAACT ACCTAACTTG AATCTTCTCA ACGTCGCCGA CCTCTTTGAT GAACTTAGGC
V E D K T K Q L R K L R G E D L D G L N L E E L Q R L E K L L E S>

420 440 460 480 500
GACTTAGCCG TGTGTCTGAA AAGAAGGGCG AGTGTGTGAT GAGCCAAATT TTCTCACTTG AGAAACGGGG ATCGGAATTG GTGGATGAGA ATAAGAGACT
CTGAATCGCG ACACAGACTT TTCTTCCCGC TCACACACTA CTCGGTTTAA AAGAGTGAAC TCTTTGCCCC TAGCCTTAAC CACCTACTCT TATTCTCTGA
G L S R V S E K K G E C V M S Q I F S L E K R G S E L V D E N K R L>

520 540 560 580 600
GAGGGATAAA CTAGAGAGCT TGGAAAGGGC AAAACTGACG ACGCTTAAAG AGGCTTTGGA GACAGAGTCG GTGACCACAA ATGTGTCAAG CTACGACAGT
CTCCCTATTT GATCTCTGCA ACCTTTCCCG TTTTGACTGC TGCGAATTC TCCGAAACCT CTGTCTCAGC CACTGGTGTG TACACAGTTC GATGTGTGCA
R D K L E T L E R A K L T T L K E A L E T E S V T T N V S S Y D S>

620 640 660
GGAACTCCCC TTGAGGATGA CTCGACACT TCCTGAAGC TTGGGCTTCC ATCTTGGGAA TGA
CCTTGAGGGG AACTCTACT GAGGCTGTGA AGGACTTCG AACCCGAAGG TAGAACCTT ACT
G T P L E D D S D T S L K L G L P S W E *>

SEQ ID NO:39 and SEQ ID NO:40

Arabidopsis AGL27 cDNA and Arabidopsis AGL27 amino acid sequence

20 40 60 80 100
ATGGGAAGAA GAAAAATCGA GATCAAGCGA ATCGAGAACA AAAGCAGTCG ACAAGTCACT TTCTCCAAAC GACGCAATGG TCTCATCGAC AAAGCTCGAC
TACCCTTCTT CTTTTATGCT CTAGTTCGCT TAGCTCTTGT TTCTGTGAGC TGTTCACTGA AAGAGGTTTG CTGCGTTACC AGAGTAGCTG TTTGAGCTG

M G R R K I E I K R I E N K S S R Q V T F S K R R N G L I D K A R>

120 140 160 180 200
AACITTCGAT TCTCTGTGAA TCCTCCGTCG CTGTTGTCGT CGTATCTGCC TCCGGAAAAC TCTATGACTC TTCTCCGGT GACGACATTT CCAAGATCAT
TTGAAAGCTA AGAGACACTT AGGAGGCAGC GACAACAGCA GCATAGACGG AGGCCTTTTC AGATACTGAG AAGGAGGCCA CTGCTGTAAA GGTTCTAGTA
Q L S I L C E S S V A V V V V S A S G K L Y D S S S G D D I S K I I>

220 240 260 280 300
TGATCGTTAT GAAATACAAC ATGCTGATGA ACTTAGAGCC TTAGATCTTG AAGAAAAAAT TCAGAATTAT CTTCCACACA AGGAGTTACT AGAAACAGTC
ACTAGCAATA CTTTATGTTG TACGACTACT TGAATCTCGG AATCTAGAAC TTCTTTTTC AGTCTTAATA GAAGGTGTGT TCCTCAATGA TCCTTGTCTAG
D R Y E I Q H A D E L R A L D L E E K I Q N Y L P H K E L L E T V>

320 340 360 380 400
CAAGCAAGC TTGAAGAACC AAATGTCGAT AATGTAAGTG TAGATTCTCT AATTTCTCTG GAGGAACAAC TTGAGACTGC TGTGTCCGTA AGTAGAGCTA
GTTTCGTTTC AACTTCTTGG TTTACAGCTA TTACATTCAC ATCTAAGAGA TTAAGAGAC CTCCTTGTTC AACTCTGACG AGACAGGCAT TCATCTCGAT
Q S K L E E P N V D N V S V D S L I S L E E Q L E T A L S V S R A>

420 440 460 480 500
GGAAGGCAGA ACTGATGATG GAGTATATCG AGTCCCTTAA AGAAAAGGAG AAATTGCTGA GAGAAGAGAA CCAGGTTCTG GCTAGCCAGC TGTCTAGAGAA
CCTTCGCTCT TGACTACTAC CTCATATAGC TCAGGGAATT TCTTTTCTC TTTAACGACT CTCTTCTCTT GGTCCAAGAC CGATCGGTGC ACAGTCTCTT
R K A E L M M E Y I E S L K E K E K L L R E E N Q V L A S Q L S E K>

520 540 560 580 600
GAAAGGTATG TCTCACCAT GAAAGATACT CAAAACCCGA TGGGAAAGAA TACGTTGCTG GCAACAGATG ATGAGAGAGG AATGTTTCCG GGAAGTAGCT
CTTTCATAC AGAGTGGCTA CTTTCTATGA GTTTTG3GCT ACCCTTTCTT ATGCAACGAC CGTTGTCTAC TACTCTCTCC TTCAAAGGC CCTTCATCGA
K G M S H R * K I L K T R W E R I R C W Q Q M M R E E C F R E V A>

620 640 660 680
CCGGCAACAA AATACCGGAG ACTCTCCCGC TGCTCAATTA GCCACCATCA TCAACGGCTG AGTTTTACCC TTAACCTCAA AGCCTGA
GGCCGTTGTT TTATGGCCTC TGAGAGGGCG ACGAGTTAAT CGGTGGTAGT AGTTGCCGAC TCAAAAGTGG AATTGTAGTT TCGACT
P A T K Y R R L S R C S I S H H H Q R L S F H L K L K A *>

SEQ ID NO:41

Arabidopsis SEP1 genomic sequence

-2981 -2961 -2941 -2921 -2901
CAGATCTCTT GGCATGTGTC GAAATGTGG AGATCTTAAG AATGTAGCTT GTGGCCGTTG CAAAGGAACA GGAACAATCA AATCAGGAGG ATTCTTTGGT
GTCTAGAGAA CCGTACACAG CTTTACACC TCTAGAATTG TTACATCGAA CACCGGCAAC GTTTCCTTGT CCTTGTAGT TTAGTCTCC TAAGAAACCA

-2881 -2861 -2841 -2821 -2801
TTCACTGACT CATCAACAC AAGATCAGTG GCTTGCATA ATTGCAAGC CAAAGGTTGT TTCCCTTGCC CTGAATGCTC AAAATCTTGA CCATTTTCTC
AAGTCACTGA GTAGTTTGTG TTCTAGTCAC CGAACGCTAT TAACGGTTTC GTTTCCARCA AAGGGAACGG GACTTACGAG TTTTAGAACT GGTAAAGAG

-2781 -2761 -2741 -2721 -2701
GGTATTTTAT AGTTGTTTCA TCTTCTGAC ACTATGATAA GTGTAATCGG TCCATTGGTA ATGTAATGT TAAAGTTGAA GAATGTCTTG TTTATTCGAG
CCATAAAATA TCAACAAAGT AGAAGAACTG TGATACTATT CACATTAGCC AGGTAACCAT TACCATTACA ATTTCAACTT CTTACAGAAC AAATAAGCTC

-2681 -2661 -2641 -2621 -2601
AAGTCTCTTA TTCCAAATCT TGATCTGTTA CTGCAAAATA GGCACCTTGC TTAGATGTAC CGGATGCTTA TGAATTACTG AGTAGGTTAA CTTTAACCGG
TTCAGAGAAT AAGGTTAAGA ACTAGACAAT GACGTTTATT CCGTGAAACG AATCTACATG GCCTACGAAT ACTTAATGAC TCATCCAATT GAAATTGGCC

-2581 -2561 -2541 -2521 -2501
GTTTATCTGT CATTAACCG GAGAAATCA TCTAGTAACC AAATGCTCTG CTGGACCTTT CTTTCAGTGA GCAACTATAG GTGGGTTTTT GGCAGTTGAT
CAAAATAGCA GTAATTTGGC CTCTTAAAGT AGATCATTGG TTTACAGAC GACCTGGAAA GAAAGTCACT CGTTGATATC CACCCAAAAA CCGTCAACTA

-2481 -2461 -2441 -2421 -2401
GTACCATAAT TGGTGCAAC ACACATTTT CTTGAATTTT TGTTTAACTT AAATAAAGTT ACTTCGTTTT CTGTTTTTT TTAATATGAA TAAAAAAAT
CATGGTATTA ACCACGTTTG TGTGTAATAA GAACCTAAAA ACAAAATGAA TTTATTTCAA TGAAGCAAAA GAACAAAAAA AATTATACCT ATTTTTTTTA

-2381 -2361 -2341 -2321 -2301
CAACCATAAC TGATAGTAGG TTGTTATCTT TTATCAAAAC AAATAAAGTT AATAGGCAGA AAAATAATGT TCTATAGAAT CAATTATGAA AATGCCATTT
GTTGGTATTG ACTATCATCC AACCAATAGA AATAGTTTTG TTTATTTCAA TTATCCGTCT TTTTATTAC AGATATCTTA GTTAATACTT TTACGTTAAA

-2281 -2261 -2241 -2221 -2201
TTTGGGATGG CATTTGTGGA TTTTGCCCTT TTTTAAATAG TTTGTAATT TTGCCATTTT TCAGGTTACG TGAATGAATA TACGTTTTAT TCATTATGTT
AAACCCATACC GTAACACCT AAAACGGGAA AAAAATTATC AAACACTTAA AACGGTAAAA AGTCCAATGC ACTTACTTAT ATGCAAAATA AGTAATACAA

-2181 -2161 -2141 -2121 -2101

TGGGTTTACT CGGTTGTGGT TGTCTTAGG GTTTAGTATT TTGTGTAAAC TACGTATTTT TACCAAAAAA AGTCGGAAT CCATATATTT TTAATCTTA
 ACCCAAATGA GCCAACACCA ACAAGATCC CAAATCATAA AACACATTG ATGCATAAAA ATGGTTTTTT TCAGGCTTTA GGTATATAAA AATTTAGAAT

-2081 -2061 -2041 -2021 -2001
 GAAATGGCT TATCCGTAG ATTTTAGTAA AAATGGCAAT TTCAAAGAT CTCTATAAAA AATGGCAAAA TCAACATAA TCCCTTGTCT ATATGGTGGT
 CTTTACCAG ATAGGCATTC TAAATCATT TTTACCGTTA AAGTTTTCTA GAGATATTTT TTACCGTTTT AGTTGTATT AGGGAACAGA TATACCACCA

-1981 -1961 -1941 -1921 -1901
 ATTTCTGCTA AAGTGACTT ATGGGTAGAT TTTTAGCTT CATAGATTCT TTGTGAAAA AAAATTACTT TGTACATTTT AGTGGAGTTA TTTAAATTTT
 TAAAGACGAT TTTCACTGAA TACCATCTA AAAAATCGAA GTATCTAAGA AACAGCTTTT TTTAATGAA ACATGTAAAA TCACCTCAAT AAATTTAAG

-1881 -1861 -1841 -1821 -1801
 CCAATTGAAC AAAACCATAT ATTGATGAAA TTCGCAATG CAATCCAAAA ATAATATGT TCCACTCTTT TGGTTAGCTT TTAATAAAG ATGCGTTTAA
 GGTAACTTG TTTTGTATA TAACACTTTT AAGCGTTTAC GTTAGTTTTT TAITTTATACA AGGTGAGAAA ACCAATCGAA AATTGATTTT TACGCAAAAT

-1781 -1761 -1741 -1721 -1701
 CTTTATGTA GTGTTGATC TTTTGGCAAT GGGGGACAAT GACTATACAA TCTAAGAGAT CATTTTAAAC AATATCATTC ATATTTCATC CTCCTCTTCA
 GAAATACATT CACCAACTAG AAAACCGTTA CCCCTGTGTA CTGATATGTT AGATTCTCTA GTAAATATGC TTATAGTAAG TATAAGTAG GAGAAGAAGT

-1681 -1661 -1641 -1621 -1601
 AATTTCACTT TCAGTAATTA ACCACGTTTC AATTGTAGTG TATCGGAGC TGTAAATATT ATCTAATTTA TGTACATAA TCATAACTGT AATCTTTATT
 TTAAAGTCAA AGTGATTAAT TGGTGCAAG TTAACATCAC ATAGCGCTCG ACATTTATAA TAGATTAAAT ACAATGTATT AGTATTGACA TTAGAATAA

-1581 -1561 -1541 -1521 -1501
 AGACAAAAAC ATATATACCT CACTGCAAC ACCTTCAAC ATGGATACT TGATTTAGGC ATACAAATAT TATTTCTCAT TTATTGATA TGACCTATAT
 TCTGTTTTTG TATATATGGA GTGACGTTTG TGAAGTTTG TACCTATTGA ACTAAATCCG TATGTTTATA ATAAAGAGTA AATAACTAT ACTGGATATA

-1481 -1461 -1441 -1421 -1401
 TATGTGGCTA TTTTATCAGT TTTAGTGTTT TTTATGATAA TTGAACCACT TAAATGTTA TCTCATTTTT CAATTTATTT TAACTGAAT TAAAGTAA
 ATACACCGAT AAAATAGTCA AATACCAAA AATACTATT AACTTGGTGA ATTTACAAAT AGAGTAAAA GTTAAATAAA ATTTGACTTA ATTTTCAIT

-1381 -1361 -1341 -1321 -1301
 GAAAGTATGA TCCAATAAGG CATCGACACA TGAACACCA TTTTAAGGTA GAAGATGCTT TTCTGGGCT TCTGAAAACA ACTAGAAAAT GATATGATAC
 CTTTCATCT AGGTATTTC GTAGCTGTGT ACCTTTGGGT AAAATTCCAT CTCTACGAA AAGACGCCGA AGACTTTTGT TGATCTTTTA CTATCTATG

-1281 -1261 -1241 -1221 -1201
 GTTGCTTTCA TTTATTGTA GTATTATTTA GTTTAATTC ACGCGCTTCA TATCCAGCTG CAAGACTACT ACACTTGCA ATTATGAGAC TCTCGTTAGA
 CAACGAAAGT AAATAACATT CATAATAAT CAAAATTAAG TGCAGCAAGT ATAGGTGAC GTTCTGATGA TGTGAACGT TAATACTCTG AGAGCAATCT

-1181 -1161 -1141 -1121 -1101
 AAATTACCAG GTATATTTA AAAACAAAA GAAGTAGAAT ATATTGGCAA TTATTGTAAG TAAGAAAAATA TGAGATTCTT GACCGAGTTG TTAACATATC
 TTTAATGGTC CATATTAAAT TTTTGTTTTT CTGATCTTA TATAACCGTT AATAAACTTC ATTTCTTTAT ACTCTAAGAA CTGGCTCAAC AATTGATAG

-1081 -1061 -1041 -1021 -1001
 AAACCAAAA GTTTTGGTTA AAAATAAGC TAGTACTATG TACATATGTT TTATGTTGAA AATATATTAA ACTGTATGTA AGAGGGAGTG TACTTTCAAT
 TTTGGTTTTT CAAACCAAT TTTTATTTC ATCATGATAC ATGTATACAA AATACAATT TTATATAATT TGACATACAT TCTCCCTCAC ATGAAAGTAA

-981 -961 -941 -921 -901
 TTAGATATAC ATTTCCAGCT AGTACGAGT CTCTATATAT AAATTTCTT AATATCGCTA AACAAATTTT ACTTTCAAGT TTGTAATGTG ATAAGTGAA
 AATCTATATG TAAAGTCTGA TCATGCTCCA GAGATATATA TTTGAAAGAA TTATAGCGAT TTGTTAAAA TGAAGTTCA AACATTACAC TATTCACITT

-881 -861 -841 -821 -801
 GACCGTATAT ACATACACAT GTTAATCAAC TGATAACCTT TGTGCTCGT GTGTCTAGTT ACTAGTCAAC CATCAACGT GCATGATGCT GTTTTCTTAA
 CTGGCATATA TGTATGTGTA CAATTAGTTG ACTATTGGAA ACACGGAGCA CACAGATCAA TGATCAGTTG GTAGTTTGCA CGTACTACGA CAAAAGAAAT

-781 -761 -741 -721 -701
 GAGTACTATT GTTGTGTTT ATATACTAA ACATAACAA TTTGCTATTA TGATATAAAC ATAGAATTTT CAAGCAATGA TATGTTTGA TGTTTGTAT
 CTCATGATAA CAACACAATA TATATTGATT TGTATTGTT AAACGATAAT ACTATATTG TATCTTAAAA GTTCGTTACT ATACAAATCT ACAAACATA

-681 -661 -641 -621 -601
 AAATATTCCA TAAATAGTAG ACACCCATAT ATACACAAAC ATGAATCTA CCTGAGGAGA AACACATAGA TGTTCAAATT AAATAATAAC CCTATAATGA
 TTTATAAGGT ATTTATCATC TGTGGGTATA TATGTGTTG TACTTAAAGT GGACTCCTCT TTGTGTATCT ACAAGTTTAA TTTATTATTG GGATATTACT

-581 -561 -541 -521 -501
 AAACCTAAA GTAAGTAATA CGAAATAAAA ATTTATCTT TAAATAACAT ATAAACATAT ATATACAAGT TTAATTGGTA ATGTATCAC AAGAGCCAAT
 TTTGAGATT CATTCATTAT GCCTTAITTT TAAATAGGAA ATTTATTGTA TAITTTGATA TATATGTTCA AATTAAACCAT TAACATAGTG TTCTCGGTTA

-481 -461 -441 -421 -401

TATTTGGTGA CTGTATCACA CGTGCTTAAA GAGAGCGTGG GAATGAAAGT AAAGAAGAAT AAAGAAGCAG AGAGATGGGC TAGAAATGAG AAAACACACC
ATAAACCACT GACATAGTGT GCAOGAATTT CTCGCGCACC CTACTTTTCA TTTCTTCTTA TTTCTTCGTC TCTCTACCCG ATCTTTACTC TTTTGTGTGG

-381 -361 -341 -321 -301
AAACCCTAAC CTCACCTCA CACATTTCTT ATCTTTTGCT CTCAATAGAT TCCATTGATT CAAAACAAA TTTTCATTAA GATTTTCACAA CCTCCACACA
TTTGGGATTG GAGTGGGAGT GTGTAAAGAA TAGAAAACGA GAGTTATCTA AGGTAACATA GTTTGTGTTT AAAAGTAATT CTAAAGTGTG GGAGGTGTGT

-281 -261 -241 -221 -201
CTTCCAAACA CAATTAAAGA GAGGAAAAAG AATCAATAAC CCTATAATA AAAAATCAGA CAAACAGAAG TTTCTCTTC TTCTTCCTTA AGCTAGTACC
GAAGGTTTGT GTTAATTTCT CTCCTTTTTC TTAGTTATTG GGATATTAT TTTTGTAGTCT GTTGTCTTC AAAGGAGAAG AAGAAGGAAT TCGATCATGG

-181 -161 -141 -121 -101
TTTTGTCTT GAAATTAGG TTAATTTCTT TTTCCAAAT ACCATCAATT CTCCAGACCA TAAAACTCA AAAAGATCAG ATCTTCTCTC TGAAAAAGAG
AAAAACAAGAA CTTTAATCCC AATTAAAGAA AAAAGGTTTA TGGTAGTTAA GAGGTCTGCT ATTTTGTAGT TTTTCTAGTC TAGAAAGGAG ACTTTTCTC

-81 -61 -41 -21 -1
ATACCCAACCT TATGTTTTTG TGTGCTGTGA TATAGATAAA CATTACATAC CCATATTGTT GTATAGACAT AAAAAGTGA AATTAAAGTA ACAAAGAGAA
TATGGGTTGA ATACAAAAC ACACAGACAT ATATCTATTT GTAATGTATG GGTATAAACA CATATCTGTA TTTTTCACCT TTAATTCAT TGTTTTCTT

20 40 60 80 100
ATGGGAAGAG GAAGAGTAGA GCTGAAGAGG ATAGAGAAACA AAATCAACAG ACAAGTAACG TTGCAAAAGC GTAGGAACGG TTTGTTGAAG AAAGCTTATG
TACCCTTCTC CTCTCATCT CGACTTCGCC TATCTCTGT TTTAGTTGTC TGTTCATTGC AAACGTTTCG CATCCTTGCC AAACAACTTC TTTCGAATAC

120 140 160 180 200
AATTGTCGT TCTCTGTGAT GCTGAAGTTG CTCTCATCAT CTCTCCAAC CGTGGAAAGC TCTATGAGTT TTGCAGCTCC TCAAAGTAAA CAACTCTCTC
TTAACAGACA AGAGACACTA CGACTTCAAC GAGAGTAGTA GAAGAGGTTG GCACCTTTCG AGATACTCAA AACGTCGAGG AGTTTCATTT GTTGAGAGAG

220 240 260 280 300
ACTCTTTATC AGTTTCTTGA TTGAGTTTTT GCTAGATCTG AGCTTAGATC TTTGTCTCAA GGACTTGTTA TATATAGATC ACACGATCTT GATTTCACG
TGAGAAATAG TCAAAGAACT AACTCAAAA CGATCTAGAC TCGAATCTAG AAACAGAGTT CCTGAACAAT ATATATCTAG TGTGCTAGAA CTAAGATGTC

320 340 360 380 400
AAGTTGAGTT AATTAGATTT CTGATTTTCA TTTTCTAGG TTTTCTTCCA ATTCTTGAAA TTTAAGATCT GGTTTTTTTG TTGCAATGA TTTAGAACTG
TTCAACTCAA TTAATCTAAA GAACATAAGT AAAAGATCCC AAAAAAGGT TAAGAACTTT AAATCTAGA CCAAAAAAC AACAGTTACT AAATCTTGAC

420 440 460 480 500
TGAATTTTGT AATCGAATAG ATTCCAAATC CTGATATGCA ATCTGAAAAG TTTTATATAA TTAATATATG TCTGTGTGAT TGGAACCTTA AAAGTTGTTT
ACTTAAACA TTAGCTTATC TAAGTTTAGT GACTATACGT TAGACTTTTC AAAATATAT TATTATATAC AGACACACTA ACCTTTGAAT TTTCAACAAG

520 540 560 580 600
ACAGATTTCT ATGAAAATTA CAAGTATCCA ACGTAGAATG ATAATATATG GTTACATGCA TTAACCAATT GTTAGTTTAT CATACTTTAT GGTGGTTAAA
TGCTAAAGA TACTTTTAAAT GTTCATAGGT TGCATCTTAC TATTATATAC CAATGTACGT AATTGGTAAA CAATCAAGTA GTATGAAATA CCACCAATTT

620 640 660 680 700
ACTTCAAACG CGTGTATATC TGTGAAGGCT TTGATTGTTT GTTTTCTTCT AAAACAATG TTTAATAGAT TTTAATTTAT ATGTTAAAT AGTTTGTCTT
TGAAGTTTGC GCACATATAG ACACCTCCGA AACTAACAAA CAAAAAGAA TTTTGTTCAC AAATTATCTA AAAATTAATA TACAATTTTA TCAAAACGAA

720 740 760 780 800
ACATGCATTC AAGAAAATAT AGCGATTAAAT TCTTTTTTTC AAATCACAAT TTGTAATCA AACGAAAACG TAAGATATTG CTTGCAAATG ATAGATTGA
TGATCGTAAG TTCTTTTATA TCGCTAATTA AGGAAAAAG TTTAGTGTTA AACACTTAGT TTGCTTTTGC ATCTATATAA GAACGTTTAC TATCCTAACT

820 840 860 880 900
ACTATTGATA TTTGTAAATA TAAATACGAA ACTTTACGTT TGAAGTTGA AACAATCAAA TCCAAATCAA CTCGTATATA ATCAGATAAA TAATGGAAAC
TGATAACTAT AAACATTTAT ATTTATGCTT TGAATGCAA ACTTTCAACT TTGTTAGTTT AGGTTTAGTT GAGCATATAT TAGTCTATTT ATTACCTTTG

920 940 960 980 1000
AATCTTCAAT TTTGATGGA GAATACTTTA AAAGTTGAAG AGCTTTTTTT TTATGGTGAT TTATAGGTTT AGATCTCCAA AGTCAAGTAT GATCTTTTTA
TTAGAAGTTA AAATACCTT CTATGAAAT TTTGAACCTC TCGAAAAAA AATACCACTA AATATCCAAA TCTAGAGGTT TCAGTTTATA CTAGAAAAAT

1020 1040 1060 1080 1100
ATAAATCTT ATTCTCTCT TTTGAGTTAT TTTGAGCATG CTCAAGACAC TTGATCGGTA CCAGAAATGC AGCTATGGAT CCATTGAAGT CAACAACAAA
TATTTGAGAA TAAGAGAGAA AAATCAATA AAAGTCGTAC GAGTCTGTG AACTAGCCAT GGTCTTTACG TCGATACCTA GGTAACCTCA GTTGTGTTT

1120 1140 1160 1180 1200
CCTGCCAAG AACTTGAGGT GTTCTTAAT CAAATACTAT TTTAGATTCC TATCATATCA TTTCAAGAAA GATCTTTTTT AAAAGTTTGT TTTGCGTAAA
GGACGGTTTC TTGAACCTCA CAAGAATTAA GTTTATGATA AAATCTAAGG ATAGTATAGT AAAGTTCTTT CTAGAAAAAA TTTTCAAAACA AAAGCACTTT

1220 1240 1260 1280 1300

TATTTACAGAA CAGCTACAGA GAATATCTGA AGCTTAAGGG TAGATATGAG AACCTTCAAC GTCAACAGAG GTACATATCT GTCTACCTCC GTATATTTAC
ATAAAGTCTT GTCGATGTCT CTTATAGACT TCGAATTCCC ATCTATACTC TTGGAAGTTG CAGTTGTCTC CATGTATAGA CAGATGGAGG CATATAAATG

1320 1340 1360 1380 1400
TCAATTCTGT ATCCATGTAG ATTCATATTT GTAGGTGTGT GTGGCTTTTG TTGGTGCAGA AATCTTCTTG GGGAGGATT AGGACCTTTG AATTCAAAGG
AGTTAAGACA TAGGTATATC TAAGTATAAA CATCCACACA CACCGAAAC AACCGCTCT TTAGAAGAAC CCCTCCTAAA TCCTGGAAC TTAAGTTTCC

1420 1440 1460 1480 1500
AGTTAGAGCA GCTTAGAGCT CAACITGGAG GCTCTCTCAA GCAAGTTCGG TCCATCAAGG TATCTTTATA CATGGAATCA ATGATTCAAA TGAGATTAAAT
TCAATCTCGT CGAATCTGCA GTTAGCTGTC CGAGAGAGTT CGTTCAAGCC AGGTAGTTCC ATAGAAATAT GTACCTTAGT TACTAAGTTT ACTCTAATTA

1520 1540 1560 1580 1600
TTGTGTGTT TAATTATAAC TACTATGGTG GTATGATGAT TGTITGCAGA CACAGTACAT GCTTGACCAG CTCTCGGATC TTCAAAATAA AGAGCAAATG
AACACAACAA ATTAATATTG ATGATACCAC CATACTACTA ACAAACGCTC GTGTATGTA CGAACTGGTC GAGAGCCTAG AAGTTTATT TCTCGTTTAC

1620 1640 1660 1680 1700
TTGCTTGAAA CCAATAGAGC TTTGGCAATG AAGGTATAAT TACAGAATAA ATGCATTTGG TGCCTTGCAG TCAATCTCTT TCACAGAGTT TAAGTTTCTA
AACGAACCTT GGTATCTCG AAACCGTTAC TTCCATATTA ATGTCTTATT TACGTAAACC ACGGAACGCT AGTTAGAGAA AGTGTCTCAA ATTCAAAGAT

1720 1740 1760 1780 1800
AACATTTTGG GAAACATCTC TAGTTTCTT GTTCTGATT ATAGTCTTTT GGTGAAATGT AAATGTTTAG CTGGATGATA TGATTGGTGT GAGAAGTCAT
TTGTAAAAAC CTTTGTAGAG ATCAAAAGAA CAAAGACTAA TATCAGAAAA CCACTTTTACA TTTACAAATC GACCTACTAT ACTAACCACA CTCTTCAGTA

1820 1840 1860 1880 1900
CATATGGGAG GAGGAGGAGG ATGGGAAGGT GGTGAACAGA ATGTTACCTA CGCGCATCAT CAAGCTCAGT CTCAGGGACT ATACCAGCCT CTTGAATGCA
GTATACCCTC CTCTCCTCC TACCCTTCCA CCACCTGTCT TACAATGGAT GCGCGTAGTA GTTCGAGTCA GAGTCCCTGA TATGTCGGA GAACCTTACGT

1920 1940 1960 1980 2000
ATCCAACCTC GCAATGGGG TAAATCCTTT GCCTTAAACA ATCATCTGCA AATCAGCTTG TGTACTTCAC TACTAAGATT GTACTTATAT AAGGTCTTTT
TAGGTTGAGA CGTTTACCCC ATTTAGGAAA CGGAATTTGT TAGTAGAGCT TTAGTCGAAC ACATGAAGTG ATGATTCTAA CATGAATATA TTCCAAGAAA

2020 2040 2060 2080 2100
AGTTACTTGG TGTAAAGAGG ATCATCAATG TGTGTGAACC TTTTAAAGTTG CTGTTTGGT GATGATGATG ATGATGACAG GTATGATAAT CCGGTATGCT
TCAATGAACC ACATTTCTCC TAGTAGTTAC ACACACTTGG AAAATTCAAC GACAAAACCA CTACTACTAC TACTACTGTC CATACTATTA GGCCATACGA

2120 2140 2160
CAGAGCAAAAT AACTGCGACA ACCCAAGCTC AGGCGCAGCA GGGAAACGGT TACATCCCGG GGTGGATGCT C
GTCTCGTTTA TTGACGCTGT TGGGTTCGAG TCCGCGTCGT CCCTTTGCCA ATGTAGGGCC CCACCTACGA G

SEQ ID NO:42

Arabidopsis SEP2 genomic sequence

-2981 -2961 -2941 -2921 -2901
ACGCTCTAAC CAACTGAGCT AATGGGCCAT TTGCGAATGG TAGTGTCTAT TTTACTTATT CGAATCTAAA TCGTCATAGG TAATTAAGAA GACATGCAAA
TGCGAGATTG GTTGACTCGA TTACCCGGTA AACGCTTACC ATCAGAGATA AAATGAATAA GCTTAGATTT AGCAGTATCC ATTAATCTTT CTGTACGTTT

-2881 -2861 -2841 -2821 -2801
GCTTAATCAA TGATGGATTG TTTGATTCTA CTTCTAGGTG CCACCATGTA CGCATTCATA AAATCATAAC CGGTCTGTTA CAAAACATAT TGCTTGAATG
CGAATTAGTT ACTACCTAAG AACTAAGAT GAAGATCCAC GGTGTTAACT GCGTAAGTAT TTTAGTATTG GCCAGCAAAAT GTTTGTATA ACGAATCTAC

-2781 -2761 -2741 -2721 -2701
ATTCTAAACA AATAATAGTT TTTTGTGAA ATTTTCAAAA CATATGTTAG GTAAGGTCAG GTTTTGCCAA TAAGCCTTAC TATATACAGT GGCAACATGT
TAAGATTGTT TTATTATCAA AAAACAACCT TAAAGTTTT GTATACAATC CATTCAGTC CAAAACGGTT ATTCGGAATG ATATATGTCA CCGTTGTACA

-2681 -2661 -2641 -2621 -2601
TTCTTCTACT TTGAGGATT TTGGTGAAT ATGAACCCA TGTGAGCATG ATACATGTGT TTCTTCTTCT ATTGAAATTT CCCCCAATGG TCATTGTCTC
AAGAAGATGA AACCTCTTAA AACCCACTTA TACTTTGGGT ACACCTGTAC TATGTACACA AAGAAGAAGA TAACCTTTAA GGGGGTTACC AGTAAACGAG

-2581 -2561 -2541 -2521 -2501
TTTGCGTTTC TGTTGCGCTT TCCGGTATCA ATCATATAT ATATATAACC TAAATGAGAC TAGACAATTT GAATCATTGT AAAAGGTATA AAGAAGAGAT
AAACGCAAGC ACAACGCGAA AGGCCATAGT TTAGTATATA TATATATTGG ATTTACTCTG ATCTGTAAAA CTTAGTAACA TTTTCCATAT TTCTTCTCTA

-2481 -2461 -2441 -2421 -2401
TATAGTCCAC AATTAACAAA GTAATAAGAC GGTAAATAT CAAACAAATT GAAAGGGTAA AAAAAAACA AGAGGGACAA GTCACTGTTA GAAAGGTGAC
ATATCAGGTG TTAATTTGTT CATTATCTG CCATTTTATA GTTTGTTTAA CTTCCTCATT TTTTCTTGT TCTCCCTGTT CAGTGACAAAT CTTCCTCACTG

-2381 -2361 -2341 -2321 -2301
TCCTCCCTTT GGGCCAGCCC CCTACCACAA AAGTCAAAGC TTACTTACTA TTCAGTCATA TATCGACACG TGTACTTGA ACCACATCAC CCATCCTATT

AGGAGGGAAA CCCGGTCGGG GGATGGTGT TTCAGTTTCG AATGAATGAT AAGTCAGTAT ATAGCTGTGC ACATGAAGCT TGGTGTAGTG GGTAGGATAA

-2281 -2261 -2241 -2221 -2201
ACGTAATTC CACTGTCTAG ACITTTTTTT TTTTCTTTT TTTTCTTTT TAACGTTTT TAGCTGTCTC TCTAAATTAC TACATACGGA CTGCTACGCT
TGCATTAAAG GTGACAGATC TGAAGAAAAA AAAAAAATA AAAATGAAAA ATTGCAAAAA ATCGACAGAG AGATTTAATG ATGTATGCCT GAACGATGCA

-2181 -2161 -2141 -2121 -2101
CACCTGAGAA GAAAGATCTT TGCTCGTAGA TTCTTTGTCT GAAGGAAAAA TATTGTGATT TAGTTATTTA CAATTGCATA ATTGTGTGTA GTAAATCCGC
GTGACTCTT CTCTCTAGAA ACGAGCATCT AAGAACAGA CTCTCTTTA ATAAACATAA ATCAATAAAT GTTAACGTAT TAACACACAT CATTTAGGCG

-2081 -2061 -2041 -2021 -2001
CAGAATGATA TTAGAGTGAT ACTGAGACGA CGAATGGTGT AACTTGTAAC ATATATACTA ATAAACACGA TTGATTAATA ATTTACTATA CAGTATATCC
GTCTTACTAT AATCTCACTA TGACTCTGCT GCTTACCACA TTGAACATTG TATATATGAT TATTGTGCT AACTAATTTT TAAATGATAT GTCATATAGG

-1981 -1961 -1941 -1921 -1901
AAAAATTAT GATTGAGAGT GTACATATAC AATAAGTAAT TAAACCTCAA AACCAACAG TTTTCTTTT TTTTGGTCAA CAATAATTAG AAATGAGAA
TTTGTAATA CTAACCTCA CATGTATATG TTATTCATTA ATTTGGAGTT TTGGTTTGTG AAAAAAATA AAAACCAAGT GTTATTAATC TTTACTCTTA

-1881 -1861 -1841 -1821 -1801
AAACTATTTA ACTTATAAAT TCTAGACCCA AAAACTCATA TTTTACCCTT CTGGTCTCA CCTAAAAGA CTTAATTTCC CAAACTCTT GCAAACTATG
TTTGATAAAT TGAATATTTA AGATCTGGGT TTTTGAGTAT AAAATGGGAA GAACCAGAGT GGATTTTTCT GAAATTAAGG GTTTTGAGAA CGTTTGTATC

-1781 -1761 -1741 -1721 -1701
GCCAAACATA GAAGATTGGA AAACAAATTT AAATCTACTT TCACTTTTAT AAAGAATAAT CAACGAACCA ATTAAGTTAA ACCTACATAT ATTGCTATGT
CGGTTTGTAT CTCTCAACCT TTTGTTTAAA TTTAGATGAA AGTGAAAAA TTTCTTATTA GTTGTCTGGT TAATTCATTT TGGATGTATA TAAGCATACA

-1681 -1661 -1641 -1621 -1601
GATCACATAT GTGTTATATT CCTCACGTTT TCTTCCATTT AGCTAATAAC CTTAATTACT TCAAGAAATC ATATATCAAC CGAAACTAG TAAATATAAT
CTAGTGATA CACAATATAA GGAGTGCAAG AGAAGGTAAA TCGATTATTG GAATTAATGA AGTTCTTTAG TATATAGTTG GCTTTGATC ATTTTATTTA

-1581 -1561 -1541 -1521 -1501
ATACATACTG AAAGCGCGCA AAATTTTATG CAATATTTTA AAATACCCTA CATCATAGTC TTAACATAAT AATCTTTCTG ATCAAAATTT ATTTTCATAA
TATGTATGAC TTTGCGCGGT TTTAAAATC GTTATAAAT TTTATGGGAT GTAGTATCAG AATTGATTAA TTAGAAAGAC TAGTTTATAA TAAAGTATT

-1481 -1461 -1441 -1421 -1401
TATTCATAAA TACTTATGGA TTACCTAAAC CAGGATACCT ATCCCTATAA ATCTGTCAAT CATCATGGAT TCATGGAGAC ATGGTCAGAT ATCCCACTGC
ATAAGTATTT ATGAATACCT AATGGATTG GTCCTATGAA TAGGGATATT TAGACAGTTA GTAGTACCTA AGTACCTCTG TACCAGTCTA TAGGGTGCAG

-1381 -1361 -1341 -1321 -1301
CAGATACAAT GTAACATATT GATATACTGC GGCTGATTAT TATTTTTTAC ATTAGAACGA GTTTAGATCC AAAACAAAAT TGGTATTCTC AAACAAAAT
GTCTATGTTA CATTGTATAA CTATATGACG CCGACTAATA ATAAAAAATG TAATCTTGCT CAATCTAGG TTTTGTTTTA ACCATAAGAG TTTGTTTTTA

-1281 -1261 -1241 -1221 -1201
TAAAAATGA ATACGAAAGT AATAGAACAA AACTTCAATG TTGTCGAATA GATAGGAAGC AATAGAAAAG CGACACGTAC ATGTCCATTT TAAGGTAGGA
ATTTTTAACT TATGCTTTCA TTATCTTGTT TTGAAGTTAC AACAGCTTAT CTATCTTTCG TTATCTTTTC GCTGTGCATG TACAGGTAAA ATTCATCTCT

-1181 -1161 -1141 -1121 -1101
GAGGCTTTTC TGCGGCTTGT GAAGTAAGAA AAAGAAAATG ATGATAGCTG CTTTCTGTTT ATTCAATGCA GAAGAAACCA ATGTTTCCCC AATCTCACGC
CTCCGAAAG ACGCCGAACA CTTCATTCTT TTCTTTTAC TACTATCGAC GAAAGCAAAG TAAAGTACGT CTCTTTTGGT TACAAAGGGG TTAGAGTGCG

-1081 -1061 -1041 -1021 -1001
GCCTCTCTCT ATCTACCACC ACTTGGACAA ATCCCTTTT CAGTATTAGT TTTTCTTTCC GGACATTGTA CATTCAAAG CATTCCAAGT GTCTAATAAA
CGGAGGAGGA TAGATGTTGG TGAACCTGTT TAGGGGAAAA GTCATATCA AAAAAAAGG CCTGTACAT GTAAGTTTTC GTAAGTTTCA CAGATTATTT

-981 -961 -941 -921 -901
CATAACTAAC CACTCCAAGA TGCAAAATCT AGCTACGAAC AAATTTTAAA CTATAGAGAT GAACCTTAAA TTCGGGCATT AATTAGTGGA ACTTGAGCTA
GTATTGATTG GTGAGGTTCT ACGTTTTAGA TCGATGCTTG TTTAAAATTT GATATCTCTA CTGAAATTT AAGCCCGTAA TTAATCACCT TGAATCGAT

-881 -861 -841 -821 -801
TTGATGAGTT TCTGACTTTT TTGAAGCTTA ATTGAGTTT ATATACACTA TATATAGGCT TGTAAATAA TGGATCAAAC AAGAAATATA TAACTACAA
AACTACTCAA AAGACTGAAA AACTTCGAAT TAACTCAAAA TATATGTGAT ATATATCCGA ACATTATTAT ACCTAGTTTG TTCTTTATAT ATTTGATGTT

-781 -761 -741 -721 -701
ATTGGGAATT AGGTTTTAAA ACGTTATCGT TCTATTTTAA TTCAGGCACC TTTAGAATAT CAAGATCCAT GCATGTTTCA ATATTTCTGT TGCAARATAA
TAACCTTTAA TCCAAATTTT TGCAATAGCA AGATAAAATT AAGTCCGTGG AAATCTTATA GTTCTAGGTA CGTACAAAGT TATAAGACA ACTGTTTATT

-681 -661 -641 -621 -601
ATAAAGATGT CTCAATATG AAGTTTGGC AACGTACGTG TAGACCTAAA AGAGTCGAAA CATTGGTATC TAAGTCATAT ATCTAGATGT ATATGGACAT

TATTCTACA GAGTTTATAC TTCAAACCCG TTGCATGCAC ATCTGAGTTT TCTCAGCTTT GTAACCATAG ATTCACTATA TAGATCTACA TATACCTGTA
-581 -561 -541 -521 -501
GGATTATATA ACTAGACAA GTTTGTITTA AAAACTTAAT TCATTTTCT TAATTAGTAG CAACAGACAA CTAAGTACTC ATGGCAAATA ATGGTGTCTG
CCTAATATAT TGATCTGTTG CAAACAAAAT TTTTGAATTA AGTAAAAGA ATTAATCATC GTTGATCGTT GATTGATGAG TACCGTTTAT TACCACAGAC
-481 -461 -441 -421 -401
CGTGGCACGC ACTTGGGAGA GAAGGTGTGA GAATGTTTTT TACTTTCTGT GTAAAAGATG GAAGAGAGAG AAAGAGTAAA GAAGTAGAGA GAGAGATATT
GCACCGTGGC TGAACCTCT CTTCACACT CTTACAAAA ATGAAAGACA CATTTCCTAC CTCTCTCTC TTTCTCATTT CTTCATCTCT CTCTCTATAA
-381 -361 -341 -321 -301
GTATCAACAA ACCCTAATGA TCTCTACCC TCACAAATTT TCTTATCTTT ATAGCTTTTA TAGATTCACA AAAACTTTTC TTCAGATTCA CAATCTCATC
CATAGTGGTT TGGGATTACT AGAGAGTGGG AGTGTTTAAA AGAATAGAAA TATCGAAAT ATCTAAGTGT TTTTGAAAAG AAGTCTAAGT TTAGAGTAGT
-281 -261 -241 -221 -201
ACAACCCCTC AAAAGTAAA AGATCTAAA GAATAACAA GAGCCCTAAT ATCAAATCAC AACCAAAAA ACCAAAGAAA GCTAATAAA GTTTTCTCTC
TGTGGGAG TTTTCTCTT TCTTAGATT CTTATTGTT CTGCGGATTA TAGTTTAGTG TTGGTTTTT TGGTTTCTTT CGATTAATTT CAAAAGAGAG
-181 -161 -141 -121 -101
TAGCTATTCC TCTTCTTTT TGTCTTGA AAAGTGGT TACTTCACC AAAAGATAAG ATCTTTCCCC AGAAAAAGCA ATACCCAAGT CATGTTTCTG
ATCGATAAGG AGAAGAAAAG AACAAAGACT TTTGATCCCA AATGAAGTGG TTTTCTATT TAGAAAGGGG TCTTTTCTGT TATGGGTCA GTACAAAGAC
-81 -61 -41 -21 -1
TGTGTCTGTA TATAGATAAA ACATTACATA CCCTAATAAG GTTACACAAA TAGCTATAAA AGAGGGGAAA TAAGATAGGG ATTTTTTGGG GTGAGGAAAG
ACACAGACAT ATATCTATTT TGAATGTAT GGGATTATC CAATGTGTTT ATCGATATTT TCTCCCTTTT ATTCTATCCC TAAAAACCC CACTCTTTTC
20 40 60 80 100
ATGGGAAGAG GAAGAGTAGA GCTCAAGAG ATAGAGAACA AATCAACAG ACAAGTGACG TTTGCTAAAC GTAGAAATGG TTTGCTGAAA AAAGCTTATG
TACCTTCTC CTCTCATCT CGAGTTCTCC TATCTCTGT TTAGTTGTC TGTTCACGCG AAACGATTG CATCTTTACC AAACGACTTT TTTCAATAC
120 140 160 180 200
AGCTTTCTGT TCTCTGCGAT GCTGAAGTCT CTCTCATGTT CTCTCCAC CGTGGCAAGC TCTACGAGTT CTGAGCACC TCCAAGTACT TCTCTTTCTT
TCGAAAGACA AGAGACGCTA CGACTTCAGA GAGAGTAGCA GAAGAGTGG GCACCGTTGC AGATGCTCAA GACGTCGTG AGGTTCTATG AGAGAAAGAA
220 240 260 280 300
TATACACTTA TTAGATCTGT GTGTAGATCT TTCATTTTT TAGTCTTGTG ATGAGTTTTA TCTTTCTTGA TTGCTTTTTT ACAAATACT TGATATATTT
ATATGTGAAT AATCTAGACA CACATCTAGA AAGTAAAAAG ATCAGAACAC TACTCAAAT AGAAAGAACT AACGAAAAAT TGTTTTATGA ACTATATATA
320 340 360 380 400
TCAGTTTCTT AATCTGATCT CTAATTAGGT TTTGATTATA GAAGAATAAT TCAGTACTTT CAAGTGATTG AATTTCGAGA TCTGATCTTA AITTATCAT
AGTCAAGAA TTAGACTAGA GATTAAATCA AAATAATAT CTTCTTATTA AGTCATGAAA GTTCACTAAC TTAAAGCTCT AGACTAGAA TAAATTAGTA
420 440 460 480 500
CATGTCAAAT TCTTAGGAT TTAATTGCAA TCTATTTTT GATTATCGG AGCTAGGAAA GTATCATAAT GATATACTAT TATTATCATG TAATTTCAAT
GTACAGTTTA AGAATCCCTA AATTAACGTT AGATAAAAA CTAAATAGCC TCGATCCTTT CATAGTATTA CTATATGATA ATAATAGTAC ATTAAGTAA
520 540 560 580 600
GTCTCTACAC GGATATATAT GTGATTAGAA CTGGTAAAG TAACTAAAG ATTCACAGTC TTCAATGAAA TTTAAAGAT CCAAGTAGA ATAATTAGTG
CAGAGATGTG CCTATATATA CACTAATCTT GAACCATTTT ATTTGATTTC TAAGTGICAG AAGTTACTTT AAATTTTCTA GGTTGCAICT TATTAATCAC
620 640 660 680 700
GTTCCATGCA TTAACGAGTC TAATTAAAGC TCATGAGAC ATTTAAGCAC CACATGAATT TAATATCTTT TTAATTAAGG GATCTTCTTT TTATAAATT
CAAGGTACGT AATTGGTCAG ATTAATTTCG AGTACGCTG TAAATTCGTG GTGTACTTAA ATTATAGAAA AATTAATTCC CTAGAGAGAA AATATTTAAA
720 740 760 780 800
TCTTTGTGA GTTTTAAAA TTTTAGTTTG TTTTAAAT TTATAGATTC TTCTTCTCT GATTGTGTT TTTTGTCTT TCAGCATGCT CAAGACACTG
AGAAAACAA CAAAAATTTT AAAATCAAC AAGTAATTTA AATATCTAAG AAGAAGAGGA CTAAACACAA AAACTAGAA AGTCGTACGA GTTCTGTGAC
820 840 860 880 900
GAAAGGTATC AGAAGTGTAG CTATGGCTCC ATTTGAAGTCA ACAACAAACC TGCTAAAGAG CTGAGGTTT AATCTCAAC ATCTCTCGA TCTTAATTAT
CTTCCATAG TCTTCACAT GATACCGAGG TAACTTCAGT TGTGTGTTGG ACGATTCTC GAACCTCAA TTAGAGGTTG TAGAGAAGCT AGAATTAATA
920 940 960 980 1000
TTATCTTTT TTAATTTTAT CTAAGAAAA TGTGTGATT TGAGACAAAA GCCCTTCAA GTTTCTTACA TAGATATTCA ATTTCTATT ATCTTCGCAA
AATAGGAAAA AATTAATAA GATTTCTTTT ACAACTTAAA ACTCTGTTTT CGGAAGTTT CAAAGAATGT ATCTATAAGT TAACAGATA TAGAAGCGTT
1020 1040 1060 1080 1100
TTTTCAGAAC AGCTACAGAG AGTACTTGAA GCTGAAAGGT AGATATGAAA ATCTGCAACG TCAGCAGAGG TATATACATT AATGTGGATG ATGATCATT

AAAAGTCTTG TCGATGTCTC TCATGAACTT CGACTTTCCA TCTATACTTT TAGACGTTGC AGTCGTCCTC ATATATGTAA TTACACCTAC TACTAGTAAA

1120 1140 1160 1180 1200
ATAAACAGCA TATATATATA TATATATATA TATATATATA GTTGTGTTTG ATCATGAAAG TGTGTTGCTG CAGAAATCTT CTTGGAGAGG ATCTTGGACC
TATTTGTCGT ATATATATAT ATATATATAT ATATATATAT CAAACATAAC TAGTACTTTC ACACAACGAC GTCTTTAGAA GAACCTCTCC TAGAACCTG3

1220 1240 1260 1280 1300
TCTGAATTCA AAGGAGCTAG AGCAGCTTGA GCGTCAACTA GACGGCTCTC TGAAGCAAGT TCGCTGCATC AAGGTGATTT ACTTCTGTAC ATACACTGAA
AGACTTAAAT TTCCTCGATC TCGTCGAACT CGCAGTTGAT CTGCCGAGAG ACTTCGTTCA AGCGACGTAG TTCCACTAAA TGAAGACATG TATGTGACTT

1320 1340 1360 1380 1400
AGATTACAC AAATCTTTCT CTATATATAG ACTGAGACAC ATGCATGAAA TGTTTTTGAT GCGTGAGGTT ATCTGAAAAA GCCTCTTCTT TTTTGACAGC
TCTAAGTGTG TTTAGAAAGA GATATATATC TGAATCTGTG TACGTACTTT ACAAAAACCTA CGCACTCCAA TAGACTTTTA CGGAGAAGAA AAAACGTCTG

1420 1440 1460 1480 1500
ACAGTATATG CTTGACCAGC TCTCTGATCT TCAAGGTAAG GAGCATATCT TGCTTGATGC CAACAGAGCT TTGTCAATGA AGGTATATGA TGATGTTTCT
TGTCATATAC GAACCTGGTC AGAGACTAGA AGTTCCTATC CTCGTATAGA ACGAACTACG GTTGTCTCGA AACAGTTACT TCCATATACT ACTACAAGA

1520 1540 1560 1580 1600
CTCTCTCTCC TCCAGTTTCT ATTTATAGAT GGAACCTTTA AATAGTCCAA TTTATATATA TGACTCTAAA TTTACATTC TTCAACTGCT ACATGTTTCT
GAGAGAGAG AGGTCAAAGA TAAATATCTA CCTTTGAAAT TTATCAGGTT AAATATATAT ACTCAGATTT AAAGTGTAAG AAGTTGACGA TGTACAAGA

1620 1640 1660 1680 1700
TTTGATTAT TTCTATGATA TCTTCAGGAA AGTTTGAAAA ATATTGTGTT TGTGTTAGCT GGAAGATATG ATCGGCGTGA GACATCACCA TATAGGAGGA
AAACATAATA AAGATACTAT AGAAGTCCTT TCAAACTTTT TATAACACAA AACAAATCGA CCTTCTATAC TAGCCGCACT CTGTAGTGGT ATATCTCTCT

1720 1740 1760 1780 1800
GGATGGGAAG GTGGTGATCA ACAGAATATT GCCTATGGAC ATCCTCAGGC TCATTCTCAG GGAATATACC AATCTCTTGA ATGTGATCCC ACTTTGCAAA
CCTACCTCTC CACCACTAGT TGTCTTATAA CGGATACCTG TAGGAGTCCG AGTAAGAGTC CCTGATATGG TTAGAGAACT TACACTAGGG TGAACCGTTT

1820 1840 1860 1880 1900
TTGGGTAAAT CAAACAACCT TTCTTGCCCT AAGACATCAA CTTAGGTTAT AAACAGTTAG CAGTTTGCTT TAAGCCCAAC ATTTGCTTTG TTTTCATAGG
AACCCTTTTA GTTTGTTGAA AAGAACGGAA TTCTGTAGTT GAATCCAATA TTTGTCAATC GTCAAACGAA ATTGGGGTGG TAACAGAAAC AAATATCTC

1920 1940 1960 1980 2000
GCTTTGGTTA AAACCTCGGT TGTGTTAGTCT AAGGATTCAG CACTTTGATG TCTGAAGTAT GGAATATCAA TATCTCAGAC TTGAAAATGT GGGTTTCTAT
CGAAACCAAT TTTGAGCACA ACAAATCAGA TTCTTAAGTC GTGAAACTAC AGACTTCATA CCTTTTAGTT ATAGAGTCTG AACTTTTACA CCCAAAGATA

2020 2040 2060 2080 2100
TGTTGACTTC GAAACTATGT TGTGTTGGTG TTGCAACAG ATATAGCCAT CCAGTGTGCT CAGAGCAAAAT GGCTGTGACG GTGCAAGGTC AGTCCCAACA
ACAACCTGAG CTTTGATACA ACAACACCAC AACGTTTGTC TATATCGGTA GGTACACAGA GTCTCGTTTA CCGACACTGC CACGTTCCAG TCAGGGTTGT

2120
AGGAAACGGC TACATCCCTG GCTGGATGCT G
TCCTTTGCCG ATGTAGGGAC CGACCTACGA C

SEQ ID NO:43

Arabidopsis SEP3 genomic sequence

-2981 -2961 -2941 -2921 -2901
GTCCCTTC CATTACGTTT TGACGTGGAC CCGTGCCGTC TATTTTATAG AGATTAATCC AACGGTTCTT ATCTTTTCTT CGACCTTCA CGACATTGCC
CAGGGGAAGG GTAATGCAGA ACTGCACCTG GGACAGGCG ATAAAAATCG TCTAATTAGG TTGCCAAGAA TAAGAAAGAA GCTGGGAAGT GCTGTACCG

-2881 -2861 -2841 -2821 -2801
TCAAAGCCGT CCGATTCTCA TCTCAGCCCC AATGACCCAC ATATATCACC AGTACTCCGC AACTTAGCTG TCGGTAGGA TTTCACGTGG CATTTATTTG
AGTTTCGGCA GCGTAAGAGT AGAGTGGGG TTACCTGGTG TATATAGTGG TCATGAGGCG TTGAATCGAC AGCAGATCTT AAAGTGCACC GTAAATAAAC

-2781 -2761 -2741 -2721 -2701
TTCTAGTTTG TAGTCAAAAC ATTGCAAGTT GATATGGTCC CCTATCGATC ACCGTCGTCT CTTTAGCTTC ACATCGAGAT TCTTCTTTCT TTCCTACGTG
AAGATCAAC ATCAGCTTTG TAACGTTCAA CTATACCAGG GGATAGCTAG TGGCAGCAGA GAAATCGAAG TGTAGCTCTA AGAAGAAGA AAGGATGCAC

-2681 -2661 -2641 -2621 -2601
TAATAGCATT TTTGATTTTG AGAATTTCTT TAGAACCGTT GGATCTCTCA TCGTTGGTTG ATCCATCCAT CCAATGGGA CCGTGTGTG CTCCATCCAG
ATTATCGTAA AAACATAAAC TCTTAAAGAA ATCTTGCGAA CCTAGAGAGT AGCAACCAAC TAGGTAGGTA GGTTTACCTT GGACACACAC GAGGTAGGTC

-2581 -2561 -2541 -2521 -2501
GGCATATGAT CCCAAAGCCA AAGAGTATT TCCAAGTCTT TCTTTCTTTT CTCTTTCTT TCTTACTAA CCTTTTTTTT TCTTATGCTT TAGACTAAGA
CCGTATACCTA GGGTTTCGGT TTTCTCATAA AGGTTACAGA AAGAAAGAAA GAAAGAAAAG AAGAATGATT GGAAGAAAAA AGAATACGAA ATCTGATTCT

-2481 -2461 -2441 -2421 -2401
 AATTTATTCG GCCATATCCA CTTTACGAA TATACTTCTT ACAAGATCTA GATTTTTTTG AGTTAATTCG GTGTATATAA CATTGGCATG GACTGCAATT
 TTAATAAGC CGGTATAGTG GAAATGCTT ATATGAAGAA TGTTCTAGAT CTAACAAAAAC TCAATTAAGC CACATATATT GTAACCGTAC CTGACGTTAA

-2381 -2361 -2341 -2321 -2301
 AAGTAATGGT AATGTGATCA TGATGCGATG TGTCGTTATC AGTAGTATAA TATTGATGGG CTACCCCTGGA AAACAAAATT ACGTGTATATA TGTACACAAT
 TTCATTACCA TTCACTAGT ACTACGCTAC ACAGCAATAG TCATCATATT ATAACCTACC GATGGGACCT TTGTGTTTAA TGCACAATAT ACATGTGTTA

-2281 -2261 -2241 -2221 -2201
 TTGGTAGAAC CGTAGAAATT AAAGTGAATA AAACCTTCTA TAATGTTCAA AATTATATGG TACAGATTAA TACGGAAGAA CATTACGCTT TTACGTAACA
 AACCATCTTG GCATCTTTAA TTGACTTAT TTGGAAGAT ATTACAAGTT TTAATATACC ATGTCTAATT ATGCCTTTTT GTAAGTGCAG AATGCATTGT

-2181 -2161 -2141 -2121 -2101
 ATTAAGTGA AAGTAAATTT ATCCCAAAAA TATTATATC ACATCATTGT TATATTCTA AGTTTTTTTA TATCTCTAAT GGTATATGTT TTACAGATTG
 TAATTCACCT TTCAATTTAA TAGGGTTTTT ATAAATATAG TGATGTAACA ATATAAAGAT TCAAAAAAAT ATAGAGATTA CCAATATACAA AATGTCTAAC

-2081 -2061 -2041 -2021 -2001
 TTTTTTGGG AATTTCTTAA AGAGACTTGA AGAATGTTTT TTTTTTATTT TCTTGAAATG TTTGACACTT GAAACCGTTT AAAAAGCTCAA ATATAGTATA
 AAAAAACCTT TTAAGAATT TCTCTGAATC TCTTACAAAA AAAAAATATA AGAAGTTTAC AAAGCTGTAA CTTTGGCAAA TTTTGTGAGT TATATCATAT

-1981 -1961 -1941 -1921 -1901
 TATCATTGTT GGTCTCATAC CTGTGAATTC ACCACATATA TTATCAATGG GGAAGATTGG AAAATTTTTG GGGGATCACA AAACGAAGGA AAGAGTACAA
 ATAGTAACAA CCAGAGTATG GAACATTAAAG TGGTGTATAT AATAGTTACC CCTTCTAAAC TTTTAAAAAC CCCCTAGTGT TTTGCTTCTT TTCTCATGTT

-1881 -1861 -1841 -1821 -1801
 AAAGAGAAGG AAAAGATAGA AGATATATGT TTTTAACTTC ATTGGTATGA CATCAATAAA TAAATAGTTG AATGTACTTT AGTTTCTCTT TTGTTTAAAT
 TTTCTCTTCC TTTTCTATCT TCTATATACA AAAATGAAG TAACCATACT GTAGTTAATT ATTTATCAAC TTACATGAAA TCAAGAGAGAA AACCAAAATTA

-1781 -1761 -1741 -1721 -1701
 GCACATCATC TCGATCAATT GTCATCATCT TACATTGAAT TATACGACCA GATCTGATAA CAAGTGAATT CGTACTTGCC CTTCCCTTTC TTCTCATACG
 CGTGTAGTAG AGCTAGTTAA CAGTAGTAGA ATGTAACCTA ATATGCTGGT CTAGACTATT GTTCACTTAA GCATGAACGG GAAGGGGAAG AAGAGTATGC

-1681 -1661 -1641 -1621 -1601
 TCCTTCTAAC TAATTTTGAT TGTAACCTAT AATTATATAA CCATATTTAA TTTTATTTTA TCTAAAACCA ATTGAAGCAA ATTTAAATAT CATTAATCTT
 AGGAAGATTG ATTAAGACTA ACATGGAATA TTAATATATT GGTATAAATT AAAATAAAAT AGATTTTGGT TAACCTCGTT TAATTTTATA GTATTTAGAA

-1581 -1561 -1541 -1521 -1501
 GAGTCCACAA TGAAGACAAAT ATATAAAACT CGTGCAAAAT TGCTTAAAT GCTTCTATGA GACCATGACC AAGTGAGATT AATAAGCGAT TCAATGTGCA
 CTCAGGGTGT ACTTCTGTTA TATATTTTGA GCACGTTTAA ACGAATTTTA CGAAGATACT CTGGTACTGG TTCACTCTAA TTATTCGCTA AGTTACACGT

-1481 -1461 -1441 -1421 -1401
 AATCAAAGA GAAAGAAGC TAATGGGTTT AAATATAACC AAACAGAATA ATAATGCTAT GTTTAGTTTT TCTAATTGAA TCATACCTTT GTGTCCATCA
 TTAGTTTTCT CTTTCTCTCG ATTACCAAAA TTTATATTGG TTTGTCTTAT TATTACGATA CAAATCAAAA AGATTAACCTT AGTATGGAAA CACAGGTAGT

-1381 -1361 -1341 -1321 -1301
 CCTACTTACC GGTGAGAATA AAGCAATTAC GTCTGCAACC AAAAAGCACT AAGACTTTTG GTCAGACATG ATCTCTAACA TCGGACGAAC CCTAAGATAA
 GGATGAATGG CAGCTCTTAT TTCGTTAATG CAGACGTTGG TTTTCTGTGA TTCTGAAAGC CAGTCTGTAC TAGAGATTGT AGCCTGCTTG GGATCTTATT

-1281 -1261 -1241 -1221 -1201
 CCAAAAAATA CTATATCTTA TATTCAAATC TCTGTTTATT TTATCCATTT ATGTTTCTTT TCTTCCCAT AATTTTTTTT GTGTCTCATC AGACTCTCTT
 GGTTTTATTI GATATAGAAT ATAAGTTTAG AGACAAATAA AATAGGTAAA TACAAAAGAA AGAAAGGGTA TTAACAAAAA CACAGAGTAG TCTGAGAGAA

-1181 -1161 -1141 -1121 -1101
 ACCAACTGA ATTTATCAAC ATGGTTTTTT TTTTGGCCAC ATCAAAATGG TGGTTTATAA AGTAGACTTA TACAAAAGAC ATTTCTGTTA ATTTCACTAA
 TGGTTTGACT TAAATAGTTG TACCAAAAAA AAAACCGGTG TAGTTTTACC ACCAAATATT TCATCTGATT ATGTTTTCTG TAAAGACAAT TAAAGTGATT

-1081 -1061 -1041 -1021 -1001
 CAAAAATAAT CTTAGCAGTA CTATAGATTG GAAAGGAAAA AGCAAACTCA GCAGTAAGAT TTATCAAAAC TAGCAGTAAG AGTTTTAGAT ATCATGAAAA
 GTTTTATTAT GAATCGTCAT GATATCTAAC CTTTCTCTTT TCGTTTAGAT CGTCATTCTA AATAGTTTTG ATCGTCATTG TCAAAATCTA TAGTACTTTT

-981 -961 -941 -921 -901
 CATCACAAAC GAGTAGTGTT TTACTTTACA TTTTAAACCA ATCACAAGGG TAGTCCGTA AGTTGGGAAA ATCGTACGAG GCTTCACTTA GTTAAGGTTA
 GTAGTGTGTT CTCATCACAA AATGAATGT AAAAATTGGT TAGTGTTCCT ATCAAGGCAT TCAACCTTTT TAGCATGCTC CGAAGTGGAT CAATTCCAAT

-881 -861 -841 -821 -801
 GGTACATGA TTCCCTGAAC TCGATTTTAT AAGTAAAAA GAAAAATTTA TAAATCAAA ATTTTATTATA TAAAAAATC AGGTGGATTI ATCAGACCTT
 CCAGTGACTT AAGGGACTTG AGCTAAAATA TTCATTTTTT CTTTTTAAAT ATTTTAGTTT TAAAAAATAT ATTTTATTAG TCCACCTAAA TAGTCTGGGA

5
10
15
20
25
30
35
40
45
50
55
60
65

-781 -761 -741 -721 -701
 ACCATCGAGA TGTGACACG TGTCCAACT CATTCAATGC CCTACTATTT TCTGTTTAGG GTTGAATCA CTCATCGCAC ACGCGCCATC TCCACCTTCC
 TGGTAGCTCT ACAGCTGTGC ACAGGTTTGA GTAAGTAACG GGATGATAAA AGACAAATCC CAACGTTAGT GAGTAGCGTG TGGCGGTAG AGGTGGAAGG

-681 -661 -641 -621 -601
 ATTATTAATC TCTCATTTTC AACATCACAC TCTTACGAAT CATACGATTT TAATATCTCT GTCTCTCTCA ACGTATTAAA TAAAAATGGT TTTAAATGTT
 TAATAATTAG AGAGTAAAG TTGTAGTGTG AGAATGCTTA GTATGCTAAA ATTATAGAGA CAGAGAGAGT TGCATAATTT ATTTTACCA AAATTTACAA

-581 -561 -541 -521 -501
 AGGGTTTTTT GTAGATTTT CAATTATTAA TCTCTATAAT TCGATGAAC AAGTAAAAA GCATCAAACT TTCTTGGCAG ATCAGATTTT TCTCTAAACT
 TCCCAAAAA CATCTAAAA GTTAATAATT AGAGATATTA AGCTACTTGA TTCATTTTTT CGTAGTTTGA AAGAACCGTC TAGTGTAAAA AGAGATTGTA

-481 -461 -441 -421 -401
 AAATATGGAC TGAATTTGAA AAATTAAACC ACTAGCTAGA ATAAAGTGTG GGTGAGAGTG GAACTCTAAT TTCTCTCCTT TACTAATTAT GTATAACAC
 TTTATACCTG ACTTTAACTT TTTAATTGGG TGATCGATCT TATTTCACAA CCACCTCTCAC CTTGAGATTG AAGAGAGGAA ATGATTATAA CATATTTGTG

-381 -361 -341 -321 -301
 AAAAATGCAC CAAATTTTTA GGTGAAAA TATCTAAGCA TGGATAGGAT AATTAACTT TTTCTTTTCA ATTTTGAAT ATTTGAATAA ATCCTATGAG
 TTTTACGTG GTTAAAAAT CCAAACTTTT ATAGATTCTG ACCTATCCCA TTAATTGTAA AAAAGAAAGT TAAACGTTA TAAACTTATT TAGGATACTC

-281 -261 -241 -221 -201
 GGTCTTTGGT ACACATAAT TGGAGGTAT ATAGTTGAGT CTGAGAGTAT ATTAGAAAGA GAATATTTC AATATGAAG CTGACATGTT TATATGTAAT
 CCAGAAACCA TGTGTTATTA ACCTCCATA TATCAACTCA GACTCTCATA TAATCTTTCT CTTATAAGT TCATTACTTC GACTGTACAA ATATACATGA

-181 -161 -141 -121 -101
 TTGAGAGAAG TGTGTTGAGA TTTGTACAAA TGTATATGTA CACTTTAAAA AGCAATATAA GATAGATAAA AAAAAATATA AGAAAAAAG AAAGAAAGAA
 AACTCTCTTC ACAACACTCT AAACATGTTT ACATATACAT GTGAATTTT TCGTATATTT CTATCTATTT TTTTATATTT TCTTTTTTCT TTTCTTTCTT

-81 -61 -41 -21 -1
 AGAAAGAAAG AGAGAGGCTC ATATATATAT AGAATTGCTT GCAAGGAAAG AGAGAGAGAG AGATTGAGAT ATCTTTTGGG AGAGAGAGAA GAAAAAGAA
 TCTTTCTTTC TCTCTCCGAG TATATATATA TCTTAACGAA CGTCTCTTTC TCTCTCTCTC TCTAACTCTA TAGAAAAACC TCTCTCTTCT CTTTCTTCTT

20 40 60 80 100
 ATGGGAAGAG GGAGAGTAGA ATTGAAGAGG ATAGAGAACA AGATCAATAG GCAAGTGACG TTTGCAAGA GAAGGAATGG TCTTTTGAAG AAAGCATACG
 TACCCTCTCT CCTCTCATCT TAACTTCTCC TATCTCTTGT TCTAGTTATC CGTTCACGTC AAACGTTTCT CTCTCTTACC AGAAAACTTC TTTCGTATGC

120 140 160 180 200
 AGCTTTTCACT TCTATGTGAT GCAGAAAGTT CTCTCATCAT CTTCTCAAAT AGAGGAAAGC TGTACGAGTT TTGCAAGTAG TCGAGGTATA TATCTACTTT
 TCGAAAGTCA AGATACACTA CGTCTTCAAC GAGAGTAGTA GAAGAGTTTA TCTCTTTTGG ACATGCTCAA AACGTCATCA AGCTCCATAT ATAGATGAAA

220 240 260 280 300
 TGTATATATA TTACTTATAA CATAACATT TTATATACAT ATTAAGTAAC ACAAAAAATGT CTTGTATGTA TGGGTCTCTC TGTGATGTGT TGTGTGTGCG
 ACATATATAT AATGAATATT GTATTGTAA AATATATGTA TAATTCATTG TGTTTTACAA GAACATACAT ACCCAGAGAG ACACATACACA ACAACACAGC

320 340 360 380 400
 TACGTACGTG TTCTATCATA TCCTTTTAAA AGAAGCAAAG AGGAAAAAAA ATTTGGGATA CCCCAATCT GTATCATTTT ATAACAAGTT TGCTTTTTTG
 ATGCATGCAC AAGATAGTAT AGGAAAATTT TCTTCGTTTC TCCTTTTTTT TAAACCCCTAT GGGGTTTAGA CATAGTAAAA TATTGTTCAA ACGAAAAAAC

420 440 460 480 500
 ATGTTCTTTT GTGTTTCTCT TTGATTTCCA TTTTGTGTTT TGATTTTTTT TCTATTCTC TTTACATCTA TCAAAGTTTT TTTTCTTATA TTTTATGTCT
 TACAAGAAAA CACAAGAGA AACTAAAGGT AAAACAAAAA ACTAAAAAAA AGATAAAGAG AAATGTAGAT AGTTTCAAAA AAAAGAATAT AAAATAACGA

520 540 560 580 600
 TATTTGTTTG TCTACTTAAT TCACATTATC TGAGAGAAGA ACAATCTATC TGATATGAAA TTAGGGTTAA TTTCTCTTGT GAGTACTCTT TAATTCACAT
 ATAAACAAC AGATGAATTA AGTGAATAG ACTCTCTTCT TGTAGATAG ACTATACTTT AATCCCAATT AAAGAGAACA CTCATGAGAA ATTAAGTGTA

620 640 660 680 700
 AAGCTTAAAG TTTCCACCTT TTGATTCTGG GGGTCGTCCA ATTCGATCAA ATCACTCAAT TTTGTTGTCA GATTGATATA AGTTCATAGG GGGATATTGT
 TTCAATTTTC AAAGGTGGA AACTAAGACC CCCAGCAGGT TAAGCTAGTT TAGTGAGTTA AAACAACAGT CTAATATAT TCAAGTATCC CCCTATAACA

720 740 760 780 800
 TTCCACGACA ATCCATTTTA GTAACCCCTA GGGGTTTCCA ATTTGGGTTT TTGAATTGAC GCTAATGTCA AATTCATCTA AAGTCCGTTG GATATGTATA
 AAGGTGCTGT TAGGTAATAA CATTGGGAAT CCCCAAGGTT TAAACCCCAA AACTTAACCTG CGATTACAGT TTAAGTAGAT TTCAGGCAAC CTATACATAT

820 840 860 880 900
 CTTGGGGATG GGATTCATCC TTTTCTCTGG GTTCTTTAGA TCTTCTCTTA AAAGACTAAC AGATTTTGTG GTAAACCCCTA GGAACAGTTT AAAAATCCCA
 GAACCCCTAC CCTAAGTAGG AAAAAGACC CAAGAAATCT AGAAGAGAAT TTTCTGATTG TCTAAACAA CATTGCGGAT CCTTTGTCAA TTTTATAGGT

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75

```

          920              940              960              980              1000
TTTTTAAAAA CATGTTTGA ACTTGATGAG TAAGATTAAAT GGAAGAAATG ATGTTTTTGT GTGGTGTGAA GCATGCTTCG GACACTGGAG AGGTACCAAA
AAAAATTTTT GTACAAAAC TGAACCTACT ATTCTAATTA CCTTCTTTAC TACAAAAACA CACCACACTT CGTACGAAGC CTGTGACCTC TCCATGTTTT

          1020              1040              1060              1080              1100
AGTGTAAC TAAGAGCA GAACCAATG TGCTTCAAG AGAGGCCCTTA GCAGTTGTAC CCAATTCTCT TCTCTTTCTT CTAATTACCT TAATTAATTA
TCACATTGAT ACCTCGTGGT CTGGGTTAC ACGGAAGTTC TCTCCGGAAT CGTCAACATG GGTAAAGAGA AGAGAAAGAA GATTAATGGA ATTAATTAAT

          1120              1140              1160              1180              1200
CTCTCAATTT TTAATTGAT TTTTAGAGTC AAATGATTAA TGTATAAAT TGTCATATAC TTCAGGAAC TGTAGCCAG CAGGAGTATC TCAAGCTTAA
GAGAGTTAAA AATGAAACTA AAAATCTCAG TTTACTAATT ACAATATTAA ACAGTATATG AAGTCTTGA ATCATCGGTC GTCCTCATAG AGTTCGAATT

          1220              1240              1260              1280              1300
GGAGCGTTAT GACGCTTAC AGAGAACCCA AAGTAAACT AATTAGCTTC TTCAGCTACC TTCAGAGAGT GTTTGTTTTT TTAGTAGAAT TTTTGTATGG
CCTCGCAATA CTGCGGAATG TCTCTTGGGT TTCCATTGTA TTAATCGAAG AAGTCGATGG AAGTCTCTCA CAAACAAAAA AATCATCTAA AAAAATACC

          1320              1340              1360              1380              1400
TTTTGATGTT GAAATAGGAA TCTGTTGGGA GAAGATCTTG GACCTCTAAG TACAAGAGAG CTTGAGTCAC TTGAGAGACA GCTTGATTCT TCCTTGAAGC
AAAACTACAA CTTTATCCTT AGACAACCTT CTCTAGAAC CTGGAGATTG ATGTTTCCTC GAAGTCAGTG AACTCTCTGT CGAACTAAGA AGGAACITCG

          1420              1440              1460              1480              1500
AGATCAGAGC TCTCAGGGA CTACTTTGTT CATCAATATC TTTATACACT GATCTATTTC CATAGTAAGA TTAATTTGG TGTTTAATTC TGCAGACACA
TCTAGTCTCG AGAGTCCCAT GATGAAACAA GTAGTTATAG AAATATGTGA CTAGATAAAG GTATCATTTCT AATTTAAACC ACAATTAAG ACCTCTGTGT

          1520              1540              1560              1580              1600
GTTTATGCTT GACCAGCTCA AGGATCTTCA GAGTAAGGTA AATAAGAAA CACTCATTCT CCTCTCTAAA TTCTCTATCT AAAAGTAATG TAACCAAGAA
CAAATACGAA CTGGTCAGAT TGCTAGAAGT CTCATTCCAT TTAATTTCTT GTGAGTAAGA GGAGAGATT AAGGAGTAGA TTTTCATTAC ATTGTTCTTT

          1620              1640              1660              1680              1700
AACACAAATA TTGGGAGCAG GAACGCATGC TGACTGAGAC AAATAAACT CTAAGACTAA GGGTAATTAA TATACATTCT CATATCACCA AATTAAATGCA
TTGTGTTTTT AAACCTCGTC CTTGCGTACG ACTGACTCTG TTTATTTTGA GATTCTGATT CCCATTAAIT ATATGTAAGA GTATAGTGGT TTAATTACGT

          1720              1740              1760              1780              1800
TCACTAAATT TGGTTATAAT GTGTGTGTGT ATATACATAT GTGACAGTTA GCTGATGGGT ATCAGATGCC ACTCCAGCTG AACCTAACC AAGAAGAGGT
AGTGATTTAA ACCAATATTA CACACACACA TATATGTATA CACTGTCAAT CGACTACCCA TAGTCTACGS TGAGGTCGAC TTGGGATTGG TTCTTCTCCA

          1820              1840              1860              1880              1900
TGATCACTAC GGTGCTCATC ATCATCAACA ACAACAACAC TCCCAAGCTT TCTTCCAGCC TTTGGAATGT GAACCAATTC TTCAGATCGG GTAACTTTAG
ACTAGTGATG CCAGCAGTAG TAGTAGTTGT TGTGTTGTG AGSGTTCTGA AGAAGGTGGG AAACCTTACA CTTGGGTAAG AAGTCTAGCC CATTGAATTC

          1920              1940              1960              1980              2000
ACTAGTATAA CCAATTTGAT TTGAGTTCTA TTATAAGCTT TTCTTAAGAA AGTATCTCAA ACTACTAAAT TTTATGGAGC AGGTATCAGG GGCAACAAGA
TGATCATATT GGTAAACTA AACTCAAGAT AATATTGGAA AAGAATCTTT TCATAGAGTT TGATGATTTA AAATACCTCG TCCATAGTCC CCGTGTGTTT

          2020              2040              2060
TGGAATGGGA GCAGGACCAA GTGTGAATAA TTACATGTTG GGTGGTTAC CTTATGACAC CAACCTTATT
ACCTTACCTT CGTCTGGTT CACACTTATT AATGTACAAC CCAACCAATG GAATACGTG GTTGAGATAA

```

SEQ ID NO:44

Arabidopsis AGL20 genomic sequence

```

          -2981              -2961              -2941              -2921              -2901
GAAAAAAGAA ACACCTAAGG AAGTGAATAT AATAGGCATA TACATATGAG GAAATGAAA ACAAAAGGAG CGAAAAATAG ATTTAACCTA AAAGAGGAAG
CTTTTTTTTT TGTGGATTTC TTCACTTATA TTATCCGTAT ATGTATACTC CTTTACTTTT TGTTTTCTC GCTTTTATC TAAATTTGAT TTCTCTCTTC

          -2881              -2861              -2841              -2821              -2801
TAAAGAGGTT ATAAGAGSTA AGAAAAGTAG GACCATATAA TAGCTATATT GTAGAATTTT ATTATTGGA GATATGGCAA TTTTGTGAG GGTCCCATGA
ATTCTCCAA TATCTCCAT TCTTTTCATC CTGGTATATT ATCGATATAA CATCTTAAA TAATAAACCT CTATACCGTT AAAAACACTC CCAGGGTACT

          -2781              -2761              -2741              -2721              -2701
AGACTAAAGT GTGGAGCACG ATTTATCTTT GTAATTATAA AAATAATAAA TATATTATTA TTGTCTCGGG ATTTTTCGAT TGAATGAGAA AAGTAAGAGG
TCTGATTTC CACCTCGTGC TAAATAGAAA CATTAAATTAT TTTATTATTT ATATAATAAT AACAGAGCCC TAAAAGCTA ACTACTCTTT TTCATTCTCC

          -2681              -2661              -2641              -2621              -2601
TGCGTTTTCG AATTATCATT GGCTAACGTT TGTACGTGAC TGTACGGACG ACGTTGATGT ATTTCTAATA TTGTACTCTT TTTTCCACC CTTATTCTC
ACGCAAAAGC TTAATAGTAA CCGATTGCAA ACATGCACTG ACATGCTGCG TGCAACTACA TAAAGATTAT AACATGAGAA AAAAGGGTGG GAATAAAGAG

          -2581              -2561              -2541              -2521              -2501
TAATTTCTGT ACATTAACCC CAAACTAATT TTACAACAC ATTGGTGTGT AATCATTTGT AAATTTTGT TATCTAAAA TACACTTTAT ATGTTATGAT
ATTAAGAACA TGTAAATGGG GTTTGATTAA AATGTTGTG TAACCACAAA TTAGTAACAC TTTAAACTA AATGATTTT ATGTGAAATA TACAATACTA

          -2481              -2461              -2441              -2421              -2401
TTTGATGAG CTTATGACTG GTAACTCAT GAGATTTCCT TATCACCATG TTGGAAGTTA CTAACCATAC ATCTTTTAAA TGCAAAATTC CATCATTCCT
AAACGTACTC GAATACTGAC CATTTGAGTA CTCTAAAGST ATAGTGTGAT AACCTTCAAT GATTGGTATG TAGAAAAATT ACGTTTAAGT GTAGTAAGGA

```

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

-2381	AGACTGCTAG	ACAGACATGT	ACTTACTTAT	ACAAGSTTTT	TCTAATCTAA	TGSCAACAAA	GAAACTTGTG	ACTAAACGCA	TACGTATCTC	TCATATAGTG	-2301
	TCTGACGATC	TGCTGTGACA	TGAATGAATA	TGTTCCAAAA	AGATTAGATT	ACCGTTGTTT	CTTTGAACAC	TGATTTCGGT	ATGCATAGAG	AGTATATCAC	
-2281	TAGACTAGAA	CTCTACGTAT	CTCTCATATA	GTCAATTTT	TAAAAAAAT	ATACCTTTGGG	ATCTCGAAGC	GAAAAITAGA	TTAGTTTATA	TGATTATGTA	-2201
	ATCTGATCTT	GAGATGCATA	GAGAGTATAT	CAGTATAAAA	ATTTTITTTA	TATGAAACCC	TAGAGCTTCG	CTTTTAATCT	AATCAAATAT	ACTAATACAT	
-2181	CAAAAAAAT	CGGATATTAC	TACCACTTAA	AAATAATTGT	AGTGGTCAAT	CATATCTAAA	ATTAATCGCA	GTGAACAAAA	ACCTGAAGCA	TAGCCTGGTT	-2101
	GTTTTTTTTA	GCCTATAATG	ATGCTGAATT	TTTATTAACA	TCACCAGTTA	GTATAGATTT	TAATTAGCGT	CACCTGTTTT	TGGACTTCGT	ATCGGACCAA	
-2081	CTATCTTACT	TTGATGTGTA	CACATTACTA	ACACGATTGT	TTTAATCTAT	AGGACGAATC	CTTTAAGTAA	TGTATAGTTG	GTTCAGTTAC	GTTAGATACT	-2001
	GATAGAATGA	AAGCTACACT	GTGTAATGAT	TGTGCTAACA	AAATTAGATA	TCCTGCTTAG	GAAATTCATT	ACATATCAAC	CAAGTCAATG	CAATCTATGA	
-1981	TTTTTTTTTG	GATTTGTCTC	AACCACTTAA	GAAGTATGTC	TATTTACTAG	TGGTATACGA	TGATGTTTCT	TTAAATCTGA	ATTGGGTCTA	CAAAATACAT	-1901
	AAAAACAAAC	CTAAACAGAG	TTGCTCAATT	CTTCACTAGC	ATAAATGATC	ACCATATGCT	ACTACAAAGA	AAATTTAGACT	TAAACCAGAT	GTTTTATGTA	
-1881	AACTAAACTT	CAAAACGGGT	TTATACITTA	TACAAACACG	AAAATATAAA	GATAGAGACA	ATTCAACCAGA	GAAGATGTGT	ATTTATATAA	AAATTTACCA	-1801
	TTGATTTGAA	GTTTGGCCCA	AATATGAAAT	ATGTTTGTGC	TTTTATATTT	CTATCTCTGT	TAAGTGGTCT	CTTCTACACA	TAAATATATT	TTTATATAGT	
-1781	TACAGATTTT	CGGACCTATC	TGTTTGATAT	TTAATATATA	TAAATACGTT	AACATATTTT	ACCAGAGAAG	ATGTGTATTT	TTCCGAATAA	TTAGTTTGTG	-1701
	ATGCTCTAAA	GCCTGGATAG	ACAAACTATA	AATTATATAT	ATTTATGCAA	TTGTATTAAG	TGCTCTCTTC	TACACATAAA	AAGCTTTATT	AATCAACAC	
-1681	TGGTCTCCTC	CCGATATATG	ATAAAAGATC	ATTAGATATC	GATTAAACAT	TTTATCTCCA	AAAAAGGATA	TTTTTTTGGT	GCCACTAGCT	AGACAGACG	-1601
	ACCAGGAGGA	GGGCTATATG	TATTTTCTAG	TAATCTATAG	CTAATTGTGA	AAATAGAGGT	TTTTTCTCTAT	AAAAAAACCA	CGGTGATCGA	TCTGTCTCG	
-1581	TTGCATAAGC	TGAATTAATTA	TTGGATTCTT	AAGTTACGTT	TTCTTTAGTA	ATCCGAGGGA	CCAAAAATAG	CAATAGCCTC	TTTAGACACG	TCGCTACTTA	-1501
	AAGCTATTGC	ACTTAATAAT	AACCTAAAGA	TTCAATGCAA	AAGAAATCAT	TAGGCTCCCT	GGTTTTTATC	GTTCACGGAG	AAATCTGTGC	AGCGATGAAT	
-1481	ACGCCATTGC	CCCATTTGCT	CTGTACTAGC	CTCCAAATAT	TTGGATTAAAT	GCTCACTTAG	GTAATGAGGA	AAITGTAGTA	TTTTGTAAAT	TGGTTTGTGC	-1401
	TGCGGTAAAC	GGGTAAACAGA	GACATGATCG	GAGGTTTATA	AACCTAATTA	CCAGTGAATC	CATTACTCTC	TTAACATCAT	AAAAACATTAC	ACCAAAACAG	
-1381	CAACTTATAA	AAACTTACAA	TTGCAAGTAA	TTAATATATC	ACATGGAGAT	GTAAGATTAT	GTCATATAAC	TAAAAACACA	ATTTAAGAAC	AACAATAAGA	-1301
	GTTGAATATT	TTTGAATGTT	AACGTTCAIT	AATTATAAAG	TGTACCTCTA	CATTCTAATA	CAGTATATTG	ATTTTGTGTT	TAAATCTCTG	TTGTTATTCT	
-1281	AACATGGAG	AAACAAGCAT	AGAAAAATATA	CAAAATCAAT	GAATTTTATC	TGTTGGGATG	GAAAGATATT	ATAAAAAATTG	ATTTAAACCA	ATATAGTTGT	-1201
	TTGTTACCTG	TTTGTTCGTA	TCTTTTATAT	GTTTAGTTTA	CTTAAATAG	ACAACCTTAC	CTTCTCTATA	TATTTTAAAC	TAATTTTGTG	TATATCAACA	
-1181	ATTACTCACA	GGTAAGAAAA	AACGATATTC	TTATTTTCCA	TATCAATTAC	AAGTGGGGGC	ATATAGGTAC	GAGAGAGTGT	TTGTGCTCAC	ATTTAAACAA	-1101
	TAATGAGTGT	CCATCTCTTT	TTGCTATAAG	AATAAAAGT	ATAGTTAATG	TTCAACCCCG	TATATCCATG	CTCTCTCACA	AACACAGGTC	TAATTTTGTG	
-1081	AAAAAAGATT	TTTGTTTAGG	GAAATTTAAT	AAAAATAATT	TGACAGGCAT	TTCCATCCAA	CTAGATATTT	ATGGGAGGGA	AAAAAGATGT	TATGTAAAAA	-1001
	TTTTTTCTAA	AAACAATCTT	CTTTAAATTA	TTTTTATTA	ACTGTCCGTA	AAGGTAGGTT	GATCTATAAA	TACCTCCCTC	TTTTCTACAC	ATACATTTTT	
-981	TGTCCATATG	TATCAAAATG	TGCTATTTTT	GCTCTTTCTT	AAGGCTTTTT	TCCAAAAATA	GTAAGGATG	AGGTTTCAAG	CGTCCATCAT	ATTTGCGACA	-901
	ACAGGTATAC	ATAGTTTTAT	ACGATAAAAA	CCAGAAAGAA	TTCCGAAAAA	AGGTTTTATT	CATTTCCTAC	TCCAAAGTTC	GCAGGTAGTA	TAAACGCTGT	
-881	CATATGACTG	ACTATTATGC	TCTCTCCTCT	TTCTTTCTCT	TATTTTATTA	TCTTTCTCCA	AGAAATAAAA	TAGAAAAAGAA	AATATATATG	GTTTCACAAA	-801
	GTATACTGAC	TGATAAATCG	AGGAGGGAGA	AAGAAAGAGA	ATAAAATAAT	AGAAAGAGGT	TCTTTATTTT	ATCTTTTCTT	TTATATATAC	CAAAGTGTIT	
-781	CACCATTACC	ATAACTACAA	CGAGAAGAGG	ATCTTTTATA	AGGAGAAAAA	CAGAGAGAGA	AGAGACGAGT	GTGTGAAGTT	TTTTTGTCTT	TTGTTTCTTT	-701
	GTGGTAATGG	TATTGATGTT	GCTCTTCTCC	TAGAAAAAAT	TCCTCTTTTC	TCTCTCTCTC	TCTCTGCTCA	CACACTTCAA	AAAAACAGAA	AACAAAGAAA	
-681	TATTACACAC	AAATAGATGA	AACGAGGAAA	GCTACTTCTT	TTGCTACTTC	CATAAAAGGG	TTCTTCTCTT	CGCAGAGAAT	CAACTTTGAT	CATCTTCTTC	-601
	ATAATGTGTG	TTTATCTACT	TTGCTCTCTT	CGATGAAGAA	AACGATGAAG	GTATTTTTC	AAGAAGGAAA	CGCTCTCTTA	GTTGAARACT	GTAGAAGAG	
-581	CTTCTCTTTT	TTTCTTCTTC	TCCCTCCAGT	AATGCTTATA	TAGTCTCCTC	CTATATCTCT	ACCTATACAT	ACACAAACCC	TTTATCTCTG	AAAGCTTCTC	-501
	GAAGAGAAAG	AAAGAAGAAG	AGGAGGTCA	TTACGAATAT	ATCAGAGGAG	GATATAGAGA	TGGATATGTA	TGTGTTTGGG	AAATAGGAGC	TTTCAAGGSA	
-481	CCTGGTTAGG	TTTTTATCAA	ACCTTTTTAG	CCAATCGGTA	AGATCTCTTC	GTCATGATCT	TTTCTTTTTT	CTTTTGTCTT	GTACTCTGAT	GGATCTATAA	-401
	GGACCAATCC	AAAAATAGTT	TGGGAAATTC	GTTTAGCCAT	TCTAGAGAGG	CAGTACTAGA	AAAGAAAAAA	GAAAAACGAA	CATGAGACTA	CCTAGATATT	
-381	ACTTATATGG	GTTTGGTTTC	ATTTGGTTTC	ATTTGATGTC	TTTGGTTTCT	TTGTCTTAAA	TCTCATGAAA	GGAGGTTGCA	TCCTTCAATT	AAACCGATAA	-301
	TGAATATACC	CAAAACAAAG	TAAACCAAGC	TAAACTACAC	AAACCAAGAA	AACAGGATTT	AGAGTACTTT	CCTCCAAAGT	AGGAAGTTAA	TTTGGCTATT	
-281	CAAAAGTTTC	CATTACAGAC	TTATAGATCA	GATACCTTTAG	ATTGTTTTCG	TTTTTGGGTA	CTTAATCTTT	CGTTGACTTC	ATCAGTCTTC	TCCACCCCAA	-201
	GTTTTCAAAG	GTAATGCTCG	AATATCTAGT	CTATGAAATC	TAACAAACG	AAAAACCCAT	GAAITAGAAA	GCAACTGAAG	TAGTCAGAAG	AGGTTGGGTT	
-181	ACAAAAAAGT	CATATTTGGA	TCATATCTTC	ATTTTITTTA	CCTACTCTCT	TTGATTCATA	TATGAAATGG	GTTGTTTAT	TTGTGTGACT	AATCTTGTGA	-101
	TGTTTTTTCA	GTATAAGCT	AGTATAGAAG	TAAAAAAAT	GGATGAGAGA	AACTAAGTAT	ATACTTTACC	CAACAAATA	CACACACTGA	TTAGAACAAT	
-81	TTGAGGTGGT	TGCACCATTT	ATCTACCGTT	TTCTTCAATT	TTTGAATAAA	TAATTTTATT	TTTTTTCTGT	GTGCAAGGGA	AAATTAACATA	AGAAGAAGAT	-1
	AACTCCACCA	ACGTGGTAAC	TAGATGGCAA	AGAAGTTTAA	AAACTTTTTT	ATTAATAATA	AAAAAAGACA	CACGTTCCTC	TTAATTGATT	TCTTCTCTTA	
20	ATGGTGAAGG	GCAAACTTCA	GATGAAGAGA	ATAGAGAATG	CAACAAGCAG	ACAAGTGACT	TTCTCCAAAA	GAAGGAATGG	TTTGTGTAAG	AAAGCCTTTG	100
	TACCACTCCC	CGTTTTGAGT	CTACTTCTCT	TATCTCTTAC	GTTGTCTGTC	TGTTCACTGA	AAGAGGTTTT	CTTCTTAC	AAACAACITC	TTTCGGTAAC	
120	AGCTCTCAGT	GCTTTGTGAT	GCTGAAGTTT	CTCTTATCAT	CTTCTCTCCT	AAAGGCAAAC	TTTATGAATT	CGCCAGCTCC	AAGTACGTTT	TTTTTGTCTT	200

TCGAGAGTCA CGAAACACTA CGACTTCAAA GAGAATAGTA GAAGAGAGGA TTCCCGTTTG AAATACTTAA GCGGTGAGG TTCATGCAAG AAAAACAGAA

220 240 260 280 300

TCTTACAAAT CATCCATAGA AAGAGAGAGA GAGAGAGATC TCATTAACCT CTCTATTGTG ATCTTAATTT TTTTGGTTT ATATATGGAT TTGATTGGCC

AGAATGTTTA GTAGGTATCT TTCTCTCTCT CTCTCTCTAG AGTAATTGGA GAGATAAACA TAGAATTAAG AAAAACCAAA TATATACCTA AACTAACCGG

320 340 360 380 400

TTTTTGAGAA TCACATCTCT TTGACGTTTG CTTTGAGAGG TGTGTTTAAA TGAGTTTCTT GGTTCCTGCA AAATTAGGGC TATTATTAAG GTAGTATCAA

AAAAACACCT AGTGTAGAGA AACTGCAAAC GAAACTCTCC ACACAAATTT ACTCAAGAA CCAAGACGCT TTTAATCCCG ATAATAATTT CATCATAGTT

420 440 460 480 500

GTACATATAC CCTCTTATTT ATTGTTTTTT TATTTCCGCT AGTATATCAT CTGTGTTTAA CATCTGTCTC TCTCTTTCTC AATTAGTTTC TCAAGTTATG

CATGTATATG GGAGAAATAA TAACAAAAAA ATAAAGGCGA TCATATAGTA GAACAAATTA GTAGACAGAG AGAGAAAGAG TTAATCAAAG AGTTCAATAC

520 540 560 580 600

ATATAAATAA AATGTGCTCT TTCGTAGCCA ATTTACACTT GTATATATAT TGATCTTCTT AGAGATCATG ATCAGATAGT ATTAATAAAA CAACTTTCAA

TATATTTATT TTACACGAGA AAGCATCGGT TAAATGTGAA CAATATATAA ACTAGAAGAA TCTCTAGTAC TAGTGTATCA TAATTATTTT GTTGAAAGTT

620 640 660 680 700

TTAGTATTCT TTTGGTTTGA ACTAATCTTT GTCTTGTAT TGCTTTAAGC AAAACATGTT GTTCTAATTT CTAAGTGTAG ATTAGGAAGT TGTTCATCAA

AATCATAAGA AAACCAAACT TGATTAGAAA CAGAACATA ACGAAATTCG TTTTGTACAA CAAGATTAAA GATTCACTAC TAATCTTCA ACAAGTAGT

720 740 760 780 800

TTCTGTATTT ATTAATCCCT CATGCTTCAT TTCATGCTCA TTCTAATTT AGTTCAATTT GTTTGAATAT TTTTCTCTGA TTTTGACATA GAAACTCAAA

AAGGACTAAA TAATTAGGGA GTACGAAGTA AAGTACGAGT AAGGATTAAA TCAAGTTAAA CAACCTTATA AACAAAGGACT AAAACTGTAT CTTTGAGTTT

820 840 860 880 900

GCTAGCTAGC CAAACCTAAA TGTGTATGTT TTTTGAGAA CAAAGAGATT TTATCTTGTA CTGTTAGGTA GTAGGGAAC CAAACTTACT TTTGATGAAT

CGATCGATCG GTTTGGATTT ACAACTAACA AAACTCTTA GTTTTCTCAA AATAGAACAT GACAATCCAT CATCCCTTTG GTTTGAATGA AAACCTCTTA

920 940 960 980 1000

CATTACTTCT GTAATGAAA ATGCCAGCTT TTGATCAGAT GTTTCAGACA TTTGGTCCAT TTGGGAAAGT ACTTCTTCTC CTCGAACCTA CTAATATATA

GTAATGAAGA CATTACTTTT TACGGTCGAA AACTAGTCTA CAAAGTCTGT AAACCAAGTA AACCCCTTCA TGAAGAAAGA GAGCTTGGAT GATTATATTT

1020 1040 1060 1080 1100

AGATAAGACC TCACATGTTT TTGATTTTTT AAAATAGGGG GAAAAAGTAC AAGACTTTTC AAGCTATGTC CTGTATTAGT TCTAGTGATA TCTTCAATAA

TCTATTCTGG AGTGTACAAA AACTAAAAGA TTTTATCCCC CTTTTTCATG TTCTGAAAAG TTCGATACAG GAACATAATC AGATCACTAT AGAAGTTATT

1120 1140 1160 1180 1200

GAAATGTTTT GAGAACACCA TTGGGATCTA AATTGATCT CTGATGATTT ACTTTAATGT TCCAATTATA TATGTTTTTG ACAGTATGCA AGATACCATA

CTTTACAAAA CTCTTGTGTT AACCTAGAT TTAAACTAGA GACTACTAAA TGAAATTACA AGGTATAAT ATACAAAAAC TGTCATACGT TCTATGTTAT

1220 1240 1260 1280 1300

GATCGTTATC TGAGGCATAC TAAGGATCGA GTCAGCACCA AACCGGTTTC TGAAGAAAAT ATGCAGGTTT ATTCTTTATG ATCTTCTTGC CTATATATCA

CTAGCAATAG ACTCCGATG ATTCTAGCT CAGTCGTGTT TTGCCCCAAG ACTTCTTTTA TACGTCCAAA TAAGAAATAC TAGAAGAACG GATATATAGT

1320 1340 1360 1380 1400

ATTCTTGCTA ATTAATACTT TTACTATATA ATATCAAGA GCGGTAATGA ATATAACCA AATATGTATA TAATCTCAAG GTCACAGGAT CAAGTCACAT

TAAGAACGAT TAATTATGAA AATGATATAT TATAGTTTCT CGCCATTACT TATATTGTTG TTATACATAT ATTAGAGTTC CAGTGTCTTA GTTCAGTGTA

1420 1440 1460 1480 1500

ATTTATAATT AGGATATATA TGTACATGCA ATAACATTTT TGTGATATA CCAACAGCAT TTGAAATATG AAGCAGCAAA CATGATGAAG AAAATTGAAC

TAAATATTAA TCCTATATAT ACATGTACGT TATTGTAAAG ACACATATAT GGTGTGCTGA AACTTTATAC TTCTGCTGTT GTACTACTTC TTTTAACTTG

1520 1540 1560 1580 1600

AACTCGAAGC TTCTAAACGG TTTGTGATAT ATACATATAT ACAAAACAT TATTCATCAC TTGTATATAT CTATTTTCATG ATGCATAGGA GAGTTTGATC

TTGAGCTTCG AAGATTTCGC AAACACTATA TATGTATATA TGTTTGTGTA ATAAGTAGTG AACATATATA GATAAAGTAG TACGTATCCT CTCAAACTAG

1620 1640 1660 1680 1700

AATTAGTGTT TTGTTTTTGT AATCAGTAAA CTCTTGGGAG AAGGCATAGG AACATGCTCA ATCGAGGAGC TGCAACAGAT TGAGCAACAG CTTGAGAAAA

TTAATCACAA AACAAAAACA TTAGTCATTT GAGAACCCTC TTCGATATCC TTGTACGAGT TAGCTCCTCG ACCTGTGCTA ACTCGTTGTC GAACTCTTTT

1720 1740 1760 1780 1800

GTGTCAAATG TATTCGAGCA AGAAAGGTAT GTGTATATAT TTATCTGTTA TATCTCCACA TTATAAGTAT TGTTCGAATC ATCTTCTGAA ACCACTGATA

CACAGTTTAC ATAAGCTCGT TCTTTCATA CACATATATA AATAGACAAT ATAGAGGTGT AATATTCTA ACAAGCTTAG TAGAAGACTT TGTGTAGTAT

1820 1840 1860 1880 1900

ATTATAACTC AATTCTCAT CTCTTTTGA CTCAAGTGT TAAGGAACAA ATTGAGCAGC TCAAGCAAAA GGTAAAGTAG TTTTATGAG TGTATATATA

TAATATTGAG TTAAGAGATA GAGAAATCT GAGTTCACAA ATTCTTGTG TAACTCGTGC AGTTCGTTTT CCATTTTCATC AAAAATACTC ACATATATTT

1920 1940 1960 1980 2000

CAGATATAAG TATGTATGCA AATTGTGTAA TATTCGAAGT AAGTAAAGCT CTGTGCTTGC CTTTTCACAA ATTGGAATCT AAAACTTTTG CAGGAGAAAG

GTCTATATTC ATACATACGT TTAACACATT ATAAGTTTCA TTCAATCCGA GAACACGAAC GAAAAATGTT TAACCTTAGA TTTTGAAAAA GTCTCTTTTC

2020 2040 2060 2080 2100

CTCTAGCTGC AGAAAAACGAG AAGCTCTCTG AAAAGGTATA ATATATCTCT ATGGGTCTCA AGTTAGGGTT GCACATTCGT TTTTATTATC GGTAAAGATA

GAGATCGAGC TCTTTTGCTC TTCGAGAGAC TTTTCCATAT TATATAAGAA TACCCAGAGT TCAATCCCAA CGTGTAAACA AAAAATAAG CCATTTCTAT

2120 2140 2160 2180 2200

AGAAAGTTGG GGTCTCTTTT GGGGGTTATT AGGTTAGGAG AGTCCTTACT AGTTTCTTCT GGTATCTTCT AATCATCAAC CTCTTTTAA TTTATGTTTG

TCTTTCAACC CCAAGAAAAA CCCCCAATAA TCCAAATCTC TCAGGAATGA TCAAAAAGAA CCAATAGAAG TTAGTAGTTG GAAGAAATTA AATACATAAC

2220 2240 2260 2280 2300

TTCTATATAT CTCTAATTT GCATCTATTA ATTTTGTGTA ATAATCTTAT TTGAATGCAG TGGGGATCTC ATGAAGGCGA AGTTTGTGTA AATAAGAAATC

AAGATATATA GAAGATTAAA CGTAGATAAT TAAACACAT TATTAAAGATA AACTTACGTC ACCCCTAGAG TACTTTCTGCT TCAACCAAGT TTATCTTATG

2320 2340 2360 2380

AAGAAAGTAC TGGAGAGGAT GATGAAGAGA GTAGCCCAAG TTCTGAAGTA GAGACGCAAT TGTTCATTGG GTTACCTTGT TCTTCAAGAA AG

TTCTTTTCATG ACCTTCTCCA CTACTCTCT CATCGGGTTC AAGACTTCAT CTCTGCGTTA ACAAGTAACC CAATGGACAA AGAAGTTCTT TC

SEQ ID NO:45

Arabidopsis AGL22 genomic sequence

-2981 -2961 -2941 -2921 -2901

TACAAGTCAT CGCCGCGTTC GTCACTTTTCA GGATCCGCGC AGAAACTGAA CCAAAATAAT ACTTATTTTA CTCGTAAGGA AAATTTGGGC CTAATAAAG

ATGTTTCAGTA GCGGCGGCAG CAGTAAAGT CCTAGGCCGC TCTTTGACTT GGTTTTATTA TGAATAAAT GAGCATTCCT TTTAAACCCG GATTATTTTC

-2881 -2861 -2841 -2821 -2801

CCCAATAATA ATAAAAAGCC CATTAGGGAC TCCGCTTTAT GATAACGGTG ACTGTAGTTT CCTTGATGTG TCAGAGAGAG TGTGTAGTGT AGGAGTGTG

GGGTTTATTAT TATTTTTCGG GTAATCCCTG AGGCGAATAA CTATTGCCAC TGACATCAAA GGAACATCAC AGTCTCTCTC ACACATCACA TCCTGACAC

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

-2781
TAGAAGAGAA GAAGCCTTAA ATGGCTAATA GGTTAGGTGC AATGTTTCAT TAGAGAGGCT TGGAACTGTT AAGGGAAAGG TCACGAGTCG TCTACTCATA
ATCTTTCTTT CTTCGGATTT TACCGATTTT CCAATCCACG TTACAAAGTA ATCTCTCCGA ACCTTGACAA TTCCCTTTCC AGTGCTCAGC AGATGAGTAT

-2681
AAAACCTCTGA CACTTTGACC AATCAAACT CAAAGACCTC ACCAGTTGTG TCACGTGCGC CTCTAAACAC TATTCAATTT CAAATATAAA TGATTCATGC
TTTTGAGACT GTGAAACTGG TTAGTTTTGA GTTCTCGGAG TGGTCAACAC AGTGCACGCG GAGATTTGTG ATAAGTTAAA GTTTATATTT ACTAAGTACG

-2581
GGTTCCAAAC GCCAATTGAT GGATGTTCTA CCAATTTTAA TCTACTTTTA CCAAAACCATG ACAAATATGA ATAAACATTA CTTGATAATA ATTTGTGAG
CCAAGGTTTG CGGTTAACTA CCTACAAGAT GGTTTAAATT AGATGAAAAT GGTTTGTGAT TGTTTATACT TATTTGTAAT GAACTATTAT TAAACACTC

-2481
TGAACAAACT TTTTTTTTTT CGAAACCAAA CCAAGCTGAA AAAAAGCTCAA CGATTTTCTT TGTTTAAAAAT ACGTTAGAAA GGAATATGTA TTATGCCGAA
ACTTGTTTGA AAAAAAATAA GCTTTGGTTT GGTTGCACTT TTTTGTAGTT GCTAAAAGAA ACAAAATTTTA TGCAATCTTT CCTTATACAT AATACGGCTT

-2381
ATAAGTAATA TCGATCAGGC CACCTCTCTT ATAGTTATTC TCCTAGCAAC TTAAACCACT AGAAGGTTTT GTTTTCTAGT GTTTTCTAAT ATACGTCATC
TATTCATTAT AGCTAGTCCG GTGAGAGAAA TATCAATAAG AGGATCGTTG AAATTTGGTGA TCTTCCAAAA CAAAAGATCA CAAAAGATTA TATGCACTAG

-2281
AAAAATTTTCA AAAAATACTA CATTTTGTGT TTAATAAATT CCATAATTC ATACTCTGTA GAACACAAAC GCAAACCATTA TTAATATTTT GTTGTCAACA
TTTTAAAGT TTTTATGAT GTAAAAACAA AATTTTGTAA GGTATTAAGG TAATGAGCAT CTGTGTTTG CGTTTGTGAT AATTATATAA CAACAGTTGT

-2181
AAAAATTTCAA ATTATAATTC AACTATAATT GCTTGATTAC CCAATTAGAT AGAAAAGAGT TAAAGAAGAA AAGAAAAGAG TTTACAGTAA ATTAACGCAA
TTTTAAAGT TAATATTAAG TTGATATAAA CGAACTAATG GGTATATCTA TCTTTTCTCA ATTTCTCTCT TTCTTTCTCT AAATGTCATT TAATTCGGTT

-2081
ACCATATTA TATTTAATTC GGTATTAATC ACATCAACCA TATGACTTTT TTACCGTTTG CAACCTTCATA ATTCATATAT TATCATAATA AATTGCGAAT
TGGTATTAT ATAAATTTGT GCATTAATTAG TGTAGTTGGT ATACTGAAAA AATGGCAAC GTTGAAGTAT TAAGTATATC ATAGTATTAT TTAAGCGTTA

-1981
AATACACAC AAGAGTTTTC TCGGAAGAGT AAATAATCT CAAATAGGGG GTGAGTGATA CGAGCCACAT GTATTCTTGA AGGGTAGATT ATGCGCAACT
TTATGTTGTG TTCTCAAAAG AGCCTTCTCA TTTATTATGA GTTTATCCCC CACTCACTAT GCTCGGTGTA CATAAGAACT TCCCATCTAA TAACGTTTGA

-1881
TGGAGTAATA AAGAGAGAAA GAATGGGTTT GTAGTAGTTG CGTGGAGTAT CTTTATTTGG GTAAAACTTT AATTAGAAA TAAAAATCTG TACGACAAAT
ACCTCATTAT TTCTCTCTT CTTACCCAAA CATCATCAAC GCACCTCATA GAAATAAACC CATTTTGAAA TTAATCTTTT ATTTTAAGAC ATGCTCTGTA

-1781
GGATCGTGT CCAATCAGAT TTCTTGTTGG TGCTTCGGGT CTGCTTTTGG GTCCCTTTGA AAAATTTTAG TGGTGACAC TTTTATTTT ACTCTGGCTC
CCTAGCACAG GGTTAGTCTA AAGAACACCG ACAAGGCCCA CAGGGAACCC CAGGGAACCC TTTTAAATC ACCAGCTGTG AAAAATAAAA TGAGACCGAG

-1681
GTGCTCTGAG GGTCCCTCTA TTCACTGTTT CTTCGTATGA AGGTATGCTT AAACATTATT TTATTTTAAA AAACCTTTTA ATTTTATTTT CTACCTTTTA
CACGGAGCTC CCAGGGAGAT AAGTGACAAA GAAGCATACT TCCATACGAA TTTGTAATAA AATAAAAAAT TTTGGGAAAT TAAATAAAAA GAATGGAAAT

-1581
ATCACGGTTT TGTAAATTGC TTTTGTAGCT ATGGAATGAT GATTGTGGCG ATTGAAATCA TATGTTTGGT TCTGTGTTG ACGTTGGTGA AGTATATGTC
TAGTGCCAAA ACATTTTAACG AAAAATCAGA TACCTTACTA CTAACACCGC TAACCTTTAG ATACAAACCA AGACAACAC TGCRAACCAT TCATATACAC

-1481
ATTTGTAATG TTGAGCTTAT GTATTAAAA GTTAAATGAT AAATAACCTC GTAAGAAAGT GATTTTCATT AAATTTTATT TTGAGTTACA TATTCATGTC
TAAACATTAC AACTCGAATA CATAATTTTA CAATTACTA TTTATTGGAG CATTTCTTCA CTAAGTAAA TTTAAATAA AACTCAATGT ATAAGTTAAC

-1381
GTTTATAAA AAAATACTTC AGTGATGATT GATACCCCA TTTGTGTGTT AATTGTTACT GGGATTGAAC AAAATTTATT TGTGCATGAC AAACCTTTCA
CAAAATATTT TTTTATGAAG TCACTACTAA CTATGGGGT AACACACACA TTAACAATGA CCTTAACCTG TTTTAAATAA ACACGTACTG TTTGAAAGGT

-1281
AATTAGTGCA TAGATTGTAA TTGTATATG GACTACATGT ATCTGAGTAG ATATGGTTCA TTAGGTTACA AACCTCTTTT TTTAAGGACA CAATTTTTCG
TTAATCACGT ATCTAACATT AACATATTAC CTGATGTACA TAGACTCATC TATACCAAGT AATCCAATGT TTGGAGAAAA AAATTCCTGT GTTAAAAAGC

-1181
ACAAGTTATA TGCCACATGA TTGACTACTA AATTTTCAA AATTATTGCA CTAATGTCTT TGAAATTAAC AAATTTATTT GTCACTTTCCG AGTTGGATTTC
TGTTCAATAT ACGGTGTACT AACTGATGAT TTAAGAGTTT TTAATAACGT GATTACAGAA ACTTTAATTG TTTAATAAAA CAGTAAGGC TCAACCTAAG

-1081
TTACAAACCA AGGCCGAACCT CACAACTTA TTTCTTTGAG TAAAAACAA ACATTGTCTC CAGAAAAATT CTGAATGTCT ATCTTCCCAA ATGTTTTPAC
AATGTTTGGT TCCGGCTTGA GTGTTTGAAT AAAGAAAGTC ATTTTGTGTT TGTAACAGGA GTCTTTTAAA GACTTTTACAG TAGAAGGGTT TACAAAAATG

-981
ATAAATAAAA ATAATATACA GTTGATATTA TTTTGTCTCT TCTGAATTTT GTTATGAGGT ACCATTACCA TATAGTACGT AGATTTACAA AAATGAAAAAT
TATTTATTTT TATTATATGT CAACTATAAT AAAACAAGAA AGACTTAAAA CAATACTTCA TGTAATGTT ATATCATGCA TCTAAATGTT TTTACTTTTA

-881
ACGTTGTAGC CTTGTAGTGT CTTCAAGTCT TCTAGTTAGT TTTTGCAGTA AATACCAACC AATTAGTTAC AAGGAGTATA AGTGAACAAA GTGAGACAAC
TGCAACATCG GAACTACAAA GAAGTCCAGA AGATCAATCA AAAACGTCTAT TTATGTTTGG TTAATCAATG TTCCTCATAT TCACCTGTTT CACTCTGTTG

-781
TCATTTTATG CTTCCCTATA AAAAGAAATT CCCCACTGAC CCAACACAC ACTTCTCTTC TCTCTCTCAT CTCATTGGAG ACTTATAAAT CCTATTACCT
AGTAAATATC GAAGGGATAT TTTTCTTTAA GGGGTGACTG GGTTTGTGTG TGAAGAGAAG AGAGAGAGTA GAGTAACCTC TGAATATTTA GGATATGGA

-681
CACCATATCC AATAACCAAC ACACACAGAC CAATATCCAA AAAAAAAGT AAAACTAAAA ATATAATATA TATCGTTTTT TTTCCAAAAA TAATCATTTA
GTGGTATAGG TTATTTGTGG TGTGTGCTG GTTATAGGTT TTTTGTGTA TTTTGTATTT TATATATATAT ATAGCAAAAG AAGGTTTTTT ATTAGTAAAT

-581
AGAAACCCCA TCATCTTGAT AGTATTATAA AATTAATAAA CCTCTCCCTG AAAATATCTC ATCCTTCACC AATCAAAACC TTCTCATGTC TTCTCTCTC
TCTTTGGGGT AGTAGAATA TCATAATATT TTAATTATTT GGAGAGGGAC TTTTATAGAG TAGGAAGTGG TTAGTTTGTG AAGAGTACAG AAGAAGAGAG

-481
CTCGACCTTT GAGGTGGAAA ATTAATATATA TTCCCTTAGC TTTTCTCTC CTTTAGTTTT CTCTCTCTTC TTGAGTTTTT TTTCTTTTGA TCCTCTCTAA
GAGCTGGAAA CTCCACCTTT TAATTTATAT AAGGGAATCG AAAAAAGAG GAAATCAAAA GAAGAAGAAG AACTCAAAAA AAGAAAAACT AGGAGAGATT

-381
TTTCTTGTG GATTATCGA CTAGATCTAA TTCTCTCAC AAAAGACTGA GTGTGTTCTT TCTTTCAAAT CTTTCAAAAA CTAGGGTTTT TACTGTCTTG
AAAGGAACAA CTAAGTAGCT GATCTAGATT AAGAAGAGTG TTTTCTGACT CACACAAGAA AGAAGTTTAA GAAAGTTTTT GATCCCAAAA ATGACAGAAC

-281
-261
-241
-221
-201

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100

AAATCATATT TATTCTTCTA AATTAGCAA AAAGAACAG ATTTACTTTC CATTTTCAGTC GTCTTGTCAC TCTCTCTCTC TTCTTTAAAG TCTCCCTTTT
 TTTAGTATAA ATAAGAGAT TTAATCGTT TTTCTTGTC TAAATGAAAG GTAAAGTCAG AGAGAGAGAG AAGAAATTC AGAGGGAAGA
 -181 -161 -141 -121 -101
 TAGCAAAAAT TCTCTCTCTC ACAAAATTTA TTTCTCTCTG CTCTCTCTCC ATCTCTCTCT TTTACTCTCT CTTTAATCAT CTCTCAITCT
 ATCGTTTTTA AGAGAGAGAG TGTTTTAAAT AAAGGAGACC GAAGAAGAAG GAGGAGGAGG TAGAGAAGAG AAATGAGAGA GAAATTAGTA GAGAGTAAGA
 -81 -61 -41 -21 -1
 TGAATCTTGA TCCATCAAAA TCAATCCCGT TCTCGAAAGA TCCATTAAAA TCAAAACCTA AGCTCTCTCT CTGCTTCTTA GGGTTTTTTT GTTCGTGTGG
 ACTTAGAACT AGGTAGTTTT AGTTAGGCCA AGAGCTTTCT AGGTAATTTT AGTTTTGAT TCGAGAGAGA GAACGAGAT CCCAAAAA CAAGCAACAC
 20 40 60 80 100
 ATGGCGAGAG AAAAGATTCA GATCAGGAAG ATCGACAACG CAACGGCGAG ACAAGTGACG TTTTCGAAAC GAAGAAGAGG GCTTTTCAAG AAAGCTGAAG
 TACCGCTCTC TTTTCTAAGT CTAGTCCTTC TAGCTGTGTC GTTGCCGCTC TGTTCAGTGC AAAAGCTTTG CTCTCTCTCC CGAAAAGTTC TTTTCGACTTC
 120 140 160 180 200
 AACTCTCCGT TCTCTGCGAC GCGATGTGCG CTCTCATCAT CTCTCTTCC ACCGGAAGAC TGTTCGAGTT CTGTAGCTCC AGGTCTTTCT TTCTCTCTCT
 TTGAGAGGCA AGAGACGCTG CGSCTACAGC GAGAGTAGTA GAAGAGAAGG TGGCTTTTGG ACAAGCTCAA GACATCGAGG TCCAGAAAGA AAGAGAGAGA
 220 240 260 280 300
 AACTTCCCTC TCTATAGATT TCTCATAACT CATCGAAGGA ATCTTGCTTA GATCCAGACA AAAAAGCTTA AAGAGTTTTT AGATGTATAT CTGATACATA
 TTGAAGGGAG AGATATCTAA AGAGATTAGA GTAGCTTCTT TAGAACAGAT CTAGGTCTGT TTTTGAAT TTCTCAAAAA TCTACATATA GACTATGTAT
 320 340 360 380 400
 GGAGTTTACT GTATCAATCT TTATAGGACC ACTAATCTT TATATAATTA AAATAGTTGT TAGAAACATT AATCATGACC ATAAATGACA TATATAAAGT
 CCTCAAATGA CATAGTTAGA AATATCTCGG TGATGTATGA ATATATTAAT TTTATCAACA ATCTTTGTAA TTAGTACTGG TATTACTGT ATATATTTCA
 420 440 460 480 500
 GTATAGTAAA ACTCTGTATT TAGATAAAAT AAGGTATCTA ACTACGGTAA TATTCAAAAA GATGTAAATC TGGATATGCA TATATGTATA TTATTAGTAT
 CATATCATTT TGAGACATAA ATCTATTTAA TTCCATAGAT TGATGCCATT ATAAGTTTTT CTACATTAG ACCTATACGT ATATACATAT AATAATCATA
 520 540 560 580 600
 ATAAATACAT GCTCTATAGT AGGTATTTGT GTCAACCATG TATAAATCTA TGTATATAGA TATTGTGTA TGAATATGTT AAGCCGCTCA TGTCTATTTT
 TATTTATGTA CGAGATATCA TCCATAAACA CAGTTGGTAC ATATTTAGAT ACATATATCT ATAACACCAT ACTATACAAA TTGCGCAGTT ACAGTATAAA
 620 640 660 680 700
 ATATAGAAAT ATGTGGGTAC CATACCATGA GGAAGTATCT ATATGTGTGG ATGTATAAAG CTTTCCCTTT GAAGAAGTAA TCTAAAAATA ATATATATAT
 TATATCTTTA TACACCCATG GTATTGTACT CCTTCATAGA TATACACACC TACATATTTT GAAAGGGAAA CTCTCTCAT TTTCTTCTAT TATATATATA
 720 740 760 780 800
 ATATATGTAT ATGTATAGAT ATGTGGAAT CTTTATTAGT GTTGGGAAAA GTCATTTAGA GAGATATTAT TGATATTAGG GATCTAAAT GACTTATCGT
 TATATACATA TACATATCTA TACAACCTTA GAAATAATCA CAACCCCTTT CAGTAAATCT CTCTATAATA ACTATAATCC CTAGATTTTA CTGAATAGCA
 820 840 860 880 900
 ATTACAGAGA TACGATTTTG GATTTTGTAC CCACGTAGTA TCAGCTCAGT TCCTATCTTC GGGGACATAC ACACCTTCAC AGATAATTGT GTATATATGT
 TAATGTCTCT ATGTCTAAAC CTAAAACTG GGTGATCAAT AGTCGAGTCA AGGATAGAAG CCCCTGTATG TGTGAAAGTG TGTATTAACA CATATATACA
 920 940 960 980 1000
 AACTGAAAC GATAGTGTGA ACATGAAATA ATGTACATGT TTGGGATTAA ATGTGTTTTG TGGATTTGGT TTGCATCTTT TGATTTTGA TTTTGGTATA
 TTGACTTTTG CTATCACAAT TGTACTTTAT TACATGTACA AACCTPAAT TACACAAAAC ACCTAAACCA AACCTAGRAA ACTAAATCT AAAACCATAT
 1020 1040 1060 1080 1100
 TTGTCGGTGT TTACATATCG ACATTGTATA TATCAACAGT ATAGTTGTTT ATAATAAGTT ATTTATTGGA ATGTGTTTAT ATTATGAAG ATGAAGGAAG
 AACAGCCACA AATGTATACG TGTAACAATT ATAGTTGTCA TATCAACAAA TATTATTCAA TAAATAACCT TACACAAATA TAATACTTCG TACTTCTCTC
 1120 1140 1160 1180 1200
 TCCTAGAGAG GCATAACTTG CAGTCAAAGA ACTTGGAGAA GCTTGATCAG CCATCTCTTG AGTTACAGGT TAGCTACATT CTCGAAACGA CCACACATTT
 AGGATCTCTC CGTATTGAAC GTCAGTTTCT TGAACCTCTT CGAACTATGC GGTAGAGAAC TCAATGTCCA ATCGATGTAA GAGCTTTGCT GGTGTGTAAA
 1220 1240 1260 1280 1300
 TCTTTCCCGA TTTCTGTAA TTTGAAAATC GAGTATTACT COGTTGAATT ACCAATATGT TTTGATTTGT TGTATTTATT GACCAAGAAT CTCTTAAAC
 AGAAGGGCT AAAGACATTG AACGTTTTAG CTCATAATGA GGCACCTTAA TGGTTATACA AAATCTAACA ACATAAATAA CTGTTCTTA GAGAATTTTG
 1320 1340 1360 1380 1400
 TTTGTATTA TAGGTACAAA ACTTTATATT ATTGCATATG ATTAATTAGA CTCGATCCAT GTAGTAGTCA TGTAGAGTAG TCTGTGTAG AGAGTTGAGC
 AAACATAATT ATCCATGTTT TGAAATATAA TAACGTATAC TAATTAATCT GAGCTAGGTA CATCATCAGT ACATCTCATC AGGACACATC TCTCACTCG
 1420 1440 1460 1480 1500
 TTTAGATCAT TATGGATATG ATTAAGAGCT TAAATCAATG TTTTATTCTG TTAGCTGTTT GAGAACAGTG ATCACGCCCG AATGAGTAAA GAAATTCGGG
 AAATCTAGTA ATACCTATAC TAATTTCTGA ATTTAGTTAC AAAATAAGAC AATCGACCAA CTCTTGTCAC TAGTGGCGGC TTACTCATTT CTTTACGCC
 1520 1540 1560 1580 1600
 ACAAGAGCCA CCGACTPAGG TACGTTATAT ATGTATATTC TATGACTTTT GAACCTAACA TCATTTTCTA ACTAATTTTT TTTTGTATCA ACCACTATCA
 TGTCTCTCGT GCTGTATCC ATGCAATATA TACATATAAG ATACTGAAA CTTGATGTAT AGTAAAGAT TGATTAAGAA AAAAAGTAGT TGTGTATAGT
 1620 1640 1660 1680 1700
 TTTTCTAACT GTGTGTTTAC ATGATCATAT ATAGGCAAAAT GAGAGAGAGG GAACCTCAAG GACTTGACAT TGAAGAGCTT CAGCAGCTAG AGAAGGCCCT
 AAAAGATTGA CACACAAATG TACTAGTATA TATCGGTTTA CTCTCTCTC CTGGAAGTTC ACTTCTCGAA GTCGTGATC TCTTCCGGGA
 1720 1740 1760 1780 1800
 TGAAACTGGT TFGACGGTGT TGATTGAAC AAAGGTTGTT AAGAAAATTA CTTGATACCA TGTATAAGTT TCTCTAAGCT TACGAGTATG CAATTTACTA
 ACTTTGACCA AACTGGGCAC ACTAAGTTTG TTTCCAACAA TTTCTTTAAT GAACATATGT ACATATTCAA AGAGATTGCA ATGCTCATAC GTTAAATGAT
 1820 1840 1860 1880 1900
 ATACGAGATG TGTTTGAGGA GTGACAAGAT TATGAGTGAG ATCAGCGAAC TTCAGAAAAA GGTAAATAAT AACCAAAATA ACGTTTATTC TTACTTTGAT
 TATGCTCTAC ACAACGCTCT CACTGTTCTA ATACTCACTC TAGTCGCTTG AAGTCTTTTT CCATTATTA TGTGTTTTAT TGCAAAATAG AATGAACATA
 1920 1940 1960 1980 2000
 GATTTCATA TTAATTTTGG CAGTTTCAAG ATCCAAATTT TTCATCTTCT TCTCTTTTTT TTTGGTGTTC AGGGAATGCA ATTGATGGAT GAGAACAAGC
 CTAAAGTTAT AATTAAGACC GTCAAAGTTC TAGGTTTTAA AAGTAGAAGA AGAGAAAAAA AAACCAACAG TCCTTTACGT TAACACCTTA CTCTTGTTCG
 2020 2040 2060 2080 2100
 GGTGAGGCA GCAAGTATGT GTCTTACCTC CTCTGTTGAT AACAAATCCC TTTCTTTTGT CTACCAITAA CGTACACACC CCTAAATTTA ATCCCAAGTT
 CCAACTCCGT CGTTCATACA CAGAATGGGA GAGACAACTA TTGTTTAGGG AAAGAAAACA GTAGGTAATT GCATGTGTGG GGAATTAAT TAGGGGTCAA
 2120
 GTCTACAACA CATATGTTTG ATCACTACTG GAGA
 CAGATGTTGT GTATACAAAC TAGTATGACA CTCT

SEQ ID NO:46

Arabidopsis AGL24 genomic sequence

```
-2981      -2961      -2941      -2921      -2901
AGACTTACAA TAACCTTCATC AAGCAACTCA TACACGAGCA CAAAGTTTTT CCTGAATGAA TCTTCATTCA GAACACCAAG ATAATCCTTA ATAACACGAG
TCTGAATGTT ATTGAAGTAG TTCGTTGAGT ATGTGCTCGT GTTTCAAAAA GGACTTACTT AGAAGTAAGT CTGTGGTTC TATTAGGAAT TATTGTGCTC

-2881      -2861      -2841      -2821      -2801
CAATCCTTTG TAGAAGCTCC AAAACAAGAG AGGGTGACAC GTTAACCTCTC GTTGTGCGAA CAAATATAG ACCAACCAACC TTGACATGGA AGTAGTTTAC
GTTAGGAAAC ATCTTCGAGG TTTTGTCTC TCCCACTGTG CAATTGAGAG CAACAGCGTT GTTTTATATC TGGTTGTTGG AACTGTACCT TCATCAAGTG

-2781      -2761      -2741      -2721      -2701
GCCATCGACA TTCTATAAGC ACAAAAAATA AGTTAGATGA AATCATTACA GCTCACAACC AAACAGAAAG TATAATACCT ACAAAGATAG GTGGCGCCTC
CGGTAGCTGT AAGATATTCG TGTTTTTTAT TCAATCTACT TTAGTAATGT CGAGTGTGCG TTTGTCTTTC ATATTATGGA TGTTCTTATC CACCGCGGAG

-2681      -2661      -2641      -2621      -2601
TGCAATTGCCA TCCTCCTTCC AGAACTTGAC TTACCGAAG AATGTCTCTG TACTTCCTTT GGGTACCTCA GCCCGGTCTG TAGCAATAAA ACGTTACACA
ACGTAACGGT AGGAGGAAGG TCTTGAACGT AAATGCCTTC TTACAGAGAC ATGAAGGAAA CCCATGGAGT CGGCCAGAC ATCGTTATTT TGCAATGTGT

-2581      -2561      -2541      -2521      -2501
TCTTGAAACT TGTATTGGAT CCAACCAAT CGTATAATCT CAAACAAT AGCTTTCTTC TACTACATTA CATACAGATA CTCGCCCAA ACTAATTGAA
AGAACTTTGA ACATAACCTA GGTGGTTTA GCATATTAGA GTTTTGTTA TCGAAGAAG ATGATGTAAT GTATGTCTAT GAGACGGGT TGAATTAACCT

-2481      -2461      -2441      -2421      -2401
TAGTTTGTCT ATATTGTAC AATCTGATT GGAATTCAG CTCACATAA TTGTCTATCG GATAAGAAAT GTTGTAGAT CAAACAGATC AATGAGCTTA
ATCAAAACGA TATAACATG TTAGACTAAA CCTTTAAGTC GAGTGTATT AAACAGTAGC CTATCTTTTA CAACCATCTA GTTGTCTAG TTACTCGAAT

-2381      -2361      -2341      -2321      -2301
GAGAAGATTT CAATGGAATA TTCTCATGAA ACAGTGACAT AAGACTCGAC TCTGAAGAGA AAAAGCAAAA CAGGAAGAAG CAGAGAGGAT CAGATCGAGA
CTCTTTTAAA GTTACCTTTT AAGAGTACTT TGTCACGTGA TTCTGAGCTG AGACTTCTCT TTTTCGTTTT GTCCTTCTTC GTCTCTCTTA GTCTAGCTCT

-2281      -2261      -2241      -2221      -2201
AAGAGAGCTT ACAGTCACGG AAGACGATGT TATCTCCTCG CTGAGATAAG ACGAAGAATT GGGAGATCAT CATCGTTCCT TATACGCGTG GATTCGCACT
TTCTCTCGAA TGTCAGTGCC TTCTGTGACA ATAGAGGAGC GACTCTATTC TGCTTCTTAA CCCTCTAGTA GTAGCAAGGA ATATCGCCAC CTAAGGCTGA

-2181      -2161      -2141      -2121      -2101
GTTTCACCGC GAGTTTGGTT AAGTCTACTG ATCGCCGATC GGTCTCTCT TTTTGTGTGT CTGGTGGTGA GGTGGTTCAC GTTTTACCAT TTGCCGCTGT
CAAAGTGGCG CTCAAACCAA TTCAGATGAC TAGCGGCTAG CCAGAGCAGA AAAACACACA GACCACCACT CCACCAAGTG CAAAATGGTA AACGGCAGCA

-2081      -2061      -2041      -2021      -2001
TATCGTGAAG CTCTTTCATG AGACGGAGGG TTCTGTGTTT TTGGAATTA TGATTCTTTG TTCTTATATG GGCCTATTTT TAAGACATCA ATATGGCCCA
ATAGCACTTC GAAGAATGAC TCTGCCTCCC AAGACACAAA AACACTTAAT ACTAAAGAAC AAGAATATAC CCGGATAAAA ATTCTGTAGT TATACCGGGT

-1981      -1961      -1941      -1921      -1901
AATTTGCAAC TTGTTATGAG TTTAAGGAAA TAAGTAGTAA GTACTATAAA TGATGGTTCC ATCTCGGAGG AGAAAAAAA AAACATTGTT TACGAGGAAG
TTAAAGCTTG AACAACTACT AAATTCCTTT ATTCATCATT CATGATATTT ACTACCAAGC TAGAGCCTCC TCTTTTTTTT TTTGTACAA ATGCTCCTTC

-1881      -1861      -1841      -1821      -1801
CAAAATGTGA GTTGATATGA AGGTACAAC ACATAATTTA TTTTGGAG TCATAACTTT GAGGATTAAG CTGACAAAGA AGTTAGTGA AGACTTTCGG
GTTTACACT CACTATATT TCCCATGTTG TGTATTAAAT AAAAACCTTC AGTTTTGAAA CTCCTAATTC GACTGTTGCT TCCAATCACT TCTGAAAGCC

-1781      -1761      -1741      -1721      -1701
GATCGAGCAA TCGGGAGATA TACATGAGCC TAGAGGGCTG ACAAGATGAC CAAGCATTC AAATGAAAGG CTTAAGATTT TTCTTTTCT AAATCAAGT
CTAGCTCGTT AGCCCTCTAT ATGTACTCGG ATCTCCGAC TGTTCTACTG GTTCGTAAGG TTTACTTTCC GAATTCTAAA AAGAAAAAGA TTTGAGTCA

-1681      -1661      -1641      -1621      -1601
AAGAAACACA AGATATATGA AAGGTAACA AGGGTCAACA ACAAGTCTAA GCTTTTAAA CGTGTAGAT GATTCTTCTT GAACACTATT ACAATTACTG
TTCTTTGTGT TCTATATACT TTCCATGTG TCCCATGTGT TGTTGAGATT CGAAAAATTT GCACAATCTA CTAAGAAGAA CTGTGTGATA TGTTAATGAC

-1581      -1561      -1541      -1521      -1501
TTTAGTTTCA CATTTATATG ACCTTGGGAG TCTCTAGCT CGTCCCAAAT ATATTTCAA CATATTACTA TAAGATCCTA AAGACCAATA ACATTGATCT
AAATCAAGT GTAAATATAC TGGAAACCTC AGAAGATCGA GCAGGGTTTA TATAAAGTT GTATAATGAT ATTCTAGGAT TTCTGGTTAT TGTAACTAGA

-1481      -1461      -1441      -1421      -1401
ACACCAAAAA CTCTCACTTT CTGATTTTGC ACTCGCTTTT TTCTCTCCA TAAACAAAAC CAAAGGCTTA CAATACTAAA TCTGTCTCAC ATTCTTAGTG
TGTGGTTTTT GAGAGTGAAA GACTAAAACG TGAGCGAAAA AAAGGAGGGT ATTTGTTTTG GTTTCGGAAT GTTATGATTT AGACAGAGTG TAAGAATCAC

-1381      -1361      -1341      -1321      -1301
CTTATTGTGT TTAGTCATAA AGAACTTAAT CTTATACAGA TTGAAGTCTT AAAGTCATCT ATATTACTTT TCACATGTAT CATTATGAGA TGGTACGTTT
```

5

10

15

20

25
30
35
40

45

50

55

60

65

GAATAAACAA AATCAGTATT TCTTGAATTA GAATATGTCT AACTTCAGAA TTTCAGTAGA TATAATGAAA AGTGTACATA GTAATACTCT ACCATGCAAA

-1281 -1261 -1241 -1221 -1201
CCCACGAATT TTATCAGTTT AGTTTAATTT TCAGTTGTAC TTTGGGAGAA AAAATTTTACA AGATACTTGT CGGCCATGAT ATCACCCCTAG AGTTACCGGA
GGGTGCTTAA AATAGTCAAA TCAAAATAAA AGTCAACATG AAACCCTCTT TTTTAAATGT TCTATGAACA GCCGCTACTA TAGTGGGATC TCAATGGCCT

-1181 -1161 -1141 -1121 -1101
GTCCGGTGAT ATATCATTTT TAATTAGGST TAAAACTTAA AAGGGTATAA ATGGCTGATC AAACCCAAAA ATAAAAGATA ATGATGACGG TGGGAGACGA
CAGGCCACTA TATAGTAAAG ATTAATCCCA ATTTTGAATT TTCCCATATT TACCGACTAG TTTGGGTTTT TATTTTCTAT TACTACTGCC ACCCTCTGCT

-1081 -1061 -1041 -1021 -1001
GTGATCTTAT CAGGTGTCGC ATCTAGCATA TATAGGTGAA AGACTATAAA AAAGACATGA AATATTTAAT AGACACAAC TTTGTAATAA ACCAAAACCA
CACTAGAATA GTCCACACGC TAGATCGTAT ATATCCACTT TCTGATAATT TTTCTGTACT TTATAAATTA TCTGTGTTGA AAACATTATT TGGTTTTGGT

-981 -961 -941 -921 -901
AAAAGGTAGA TGAAGTGATG AACAGCATCT TCTAATTACG AATAAAAAAA GTAACCAAAC TTTCTTTCCA TTAGAATTGG TACGTAGTTC CTTGTGTATT
TTTTCCATCT ACTTGACTAC TTGTCGTAGA AGATTAATGC TTATTTTTTT CATTGGTTTG AAAGAAAAGT AATCTTAACC ATGCATCAAG GAACACATAA

-881 -861 -841 -821 -801
GTGATTTCTT TCATTTTCCA ATTATGTTTT TTTATTTTAT CATGTTACAT TTTTGATAGT GGGTAACTTT TGTATCATTT TATTTGACCT AGCCATATAT
CACTAAGAAA AGTAAAGGT TAATACAAAA AAATAAATA GTACAATGTA AAAACTATCA CCCATTGAAA ACATAGTAAA ATAACTGGA TCGGTATATA

-781 -761 -741 -721 -701
AAATCTATTA ACTTATACGG AGTAGTATTT CACGTCATTT ATTTTTATTT TGTTTTAGA TGGGAAGTTA TTCAAAACTA GACTAAAACA GTAAAACTAG
TTTAGATAAT TGAATATGCC TCATCATAAA GTGCAGTAAA TAAAAATAA ACAAATCT ACCCTTCAAT AAGTTTGGAT CTGATTTTGT CATTTTGATC

-681 -661 -641 -621 -601
GAAACCCGCT ACTGAATAAA GTTACAATTC CACATTAATC CATGACAGAC TAATTGAATT AGAAGGTTAG GTAAATTATT AAATCATAC TGTAGCAGTC
CTTTGGGCGA TGACTTATTT CAATGTTAAG GTGTAATAAG GTACTGTCTG ATTAACFTAA TCTTCCAATC CATTTAATAA TTTAGTATTG ACATCGTCAG

-581 -561 -541 -521 -501
TCTTCGTCG GCAGCTCAGT CAGACAAAAC ACAAGGTGTG TTTATGTGTT ATTTTAAATG ATTATAGTTT GGGAAAAAGA CATAATCAAA AGGGATACAA
AGAAGCAGAC CGTCGAGTCA GTCTGTTTTG TGTTCACAC AAATACACAA TAAAAATTAC TAATATCAAA CCCTTTTCT GTATTAGTTT TCCCTATGTT

-481 -461 -441 -421 -401
AACATATGCC CCATTGATAA GTATAGATCA CTGTTAGCT AAAAAAGCA GACTCTTTT TCCATCTTG AACACAAACA CAGTCACCAT CTCTCTCTCT
TTGTATACCG GGTAACATTT CATATCTAGT GACAAATCGA TTTTTTCTGT CTGAGAAAAA AGGTTAGAAC TTGTGTTTGT GTCAGTGGTA GAGAGAGAGA

-381 -361 -341 -321 -301
CTTTCTCTCT CACTCACACA TTAGGGAGTA AACAGCTACC AGAAAAACCT TTTTATCTTT CTCACAAATT TAATAAGTG GGTGCTGAGA TTGAATAACG
GAAAGAGAGA GTGAGTGTGT AATCCCTCAT TTGTCGATGG TCTTTTGGGA AAAAATAGAA GAGTGTTTAA ATTATTTCAC CCAGCTCTCT AACTTATTGC

-281 -261 -241 -221 -201
TAATCCAAGA TCCTCCAAT CACAGAAAGG TAAAGCTGT GAATCTGTGT TCTTCTTCT TAAGCAAAGT GTTTGATGAA TTCATCTAGT CCTGTCCATT
ATTAGGTTCT AGGAGGTTGA GTGCTTTTCC ATTTTCGACA CTTAGACACA AGAAGAAGA ATTCGTTTCA CAACTACTT AAGTAGATCA GGACAGGTAA

-181 -161 -141 -121 -101
CTTTGCTTCT TCATGGTTTA TGGATCTGAT CTCTCTTCT CTCTCTCTCT AGCCATTAGG GTTTCCTAAG AATATTATAT AAACCTCTCT TAGCTAACAC
GAAAACGAAG AGTACCAAT ACCTAGACTA GAGAGAAAGA GAGAGAGAGA TCGGTAATCC CAAAGGATTC TTATAATATA TTTGAGAGAA ATCGATTGTG

-81 -61 -41 -21 -1
CGTTCCAATT GGTTCCTTTC TTTGTTCTTG GTCTAAAATC TAAATGGTGT TATGGGTATA GGCAGATTCA AGAACAGTAG TGAAGGAGAG ATCTGGTAAA
GCAAGGTAA CCAAGAAAG AAACAAGAAC CAGATTTTAG ATTTACCACA ATACCCATAT CGTCTAAGT TCTTGTCTAT ACTTCTCTCT TAGACCATTT

20 40 60 80 100
ATGGCGAGAG AGAAGATAAG GATAAAGAAG ATTGATAACA TAACAGCGAG ACAAGTTACT TTCTCAAAGA GAAGAAGAGG AATCTTCAAG AAAGCCGATG
TACCGCTCTC TCTTCTATTC CTATTTCTTC TAATATTGT ATTGTGCTC TGTTCATGA AAGAGTTTCT CTCTTCTCC TTAGAAGTTC TTTCCGCTAC

120 140 160 180 200
AACTTTTCACT TCTTTGCGAT GCTGATGTTG CTCTCATCAT CTCTCTGCC ACCGGAAAGC TCTTCAGATT CTCCAGCTCA AGGTATATTC TATCTTTTTG
TTGAAAGTCA AGAAACGCTA CGACTACAAC GAGAGTAGTA GAAGAGACGG TGGCCTTTCC AGAAGCTCAA GAGTCGAGT TCCATATAAG ATAGAAAAAC

220 240 260 280 300
TTAGTAGTTG TCTTATTTTT TTCAATCCAT GTTTGTGTTT TTGAGAATAT GGTGGGATAA ATATATTAAG ATATGTATTT AAATGAGATT TTTATTTTTCT
AATCATCAAC AGAATAAAAA AAGTTAGTGA CAAACACAAA AACTCTTATA CCAACCTATT TATATAATTC TATACATAAA TTTACTCTAA AAATAAAAAA

320 340 360 380 400
CGTTTACTCT CTAAAGTTAA TTATCAGTAG GCTCGGAGAT CTCATGTACG GCATAATTTG ATGACCTAAA TTATTATACT TTAAGTATA GGATTGATGT

GCAAATGAGA GATTTC AATT AATAGTCATC CGAGCCTCTA GAGTACATGC CGTATTAAAC TACTGGATTT AATAATATGA AATTCATAT CCTAACTACA
 420 440 460 480 500
 TTTATTACTT TTTATGTATA CACATCATGT ATTTAATTCC GTTTAACATA ATATGGGTTT TTAACGTGTA ATTTTTC AATTTCATT TAGACTCATG
 5 AAATAATGAA AATACATATT GTGTAGTACA TAAATTAAGG CAAATTGTAT TATACCCAAA AATTGCACAT TAAAAAGTTA GTAAAAGTAA ATCTGAGTAC
 520 540 560 580 600
 GTTAAGATTT CTGTACTGGG AAATAAGAGA GCAGAATATT ATAGTGTGAT TTTGTGTAAT TAGGAAAGCA TATGTATATA TGGATACATA GTACTTACCA
 10 CAATTCTAAA GACATGACCC TTTATTCTCT CGTCTTATAA TATCACACTA AAAACAATTA ATCTTTTCGT ATACATATAT ACCTATGTAT CATGAATGGT
 620 640 660 680 700
 CAATTAGAAT GAATTTCTTT TCCCTTTTTT CATTTGACTT TGTGTATTAC AAAAGTCTTT GACACTGTCA CTGTGTATGA TTGGGGATTA ATCTTTAACC
 GTTAATCTTA CTTAAGAAAA AGGGAAAAAA GTAACTGAA ACACATAATG TTTTCAGAAA CTGTGACAGT GAACCATACT AACCCCTAAT TAAGAATTGG
 720 740 760 780 800
 ACTCGTTTAG TTTATCTTGG GAAGCATTAC CATAATTGGG AAACGAGTCA TCTGTCTGTA TCGTGTATGG TACTTCTGAT TACTTTTCTT TTATTATAAC
 TGAGCAAAATC AAATAGAACC CTTCGTAATG GTATTAACCC TTTGCTCAGT AGACAGACAT AGCACTACCG ATGAAGACTA ATGAAAAGAA AATAATATTG
 820 840 860 880 900
 CAAAAGGCT TCTAATGTAC TTAATTAATT TTACAATGT AATATGGACG AAGGAAATGT TTATAAGAAA GATGGATTGT TTGTGAAAC GTGTAGAATG
 GTTTTCCGA AGATTAGATG AATTAATTAA AATGTTTACA TTATACCTGC TTCCCTTACA AATATTCTTT CTACCTAACA AACCACTTIG CACATCTTAC
 920 940 960 980 1000
 AGAGACATAT TGGGAAGGTA TAGTCTTCAT GCAAGTAACA TCAACAAATT GATGGATCCA CCTTCTACTC ATCTCCGGGT ATTTTCGATA TCACCTTACTC
 TCTCTGTATA ACCCTTCCAT ATCAGAAGTA CGTTCATTGT AGTTGTTTAA CTACCTAGGT GGAAGATGAG TAGAGGCCCA TAAAGCTAT AGTGAATGAG
 1020 1040 1060 1080 1100
 TTTTTTTTTT TTGTGGATTT TAACTCTCT GCTCTTTTTA CCAAAACCTT CTCTTTTAT CAAACCTTTC TCCTATAAT ATTATCGAT GTTCACCTTG
 AAAAAAATA AACACCTAAA ATTTGAGAGA CGAGAAAAAT GGTTTGGGAA GAGAAAAATA GTTTGGGAG AGAGATATTA TAATAGGCTA CAAGTGAAC
 1120 1140 1160 1180 1200
 TTACACGTGT TTGTATAAT TTTAGCTGT AAGTCTAAAT ATAGAAACAT TGAGTGGCAT ATAATCATT ATCTTGAAGC ATCTAATTAA TTGTTTITAC
 AATGTGCACA AACAAATTA AAAATCGACA TTCAGATTTA TATCTTTGTA ACTCACCCTA TATTAGTAAT TAGAACTTCG TAGATTAATT AACCAAAATG
 1220 1240 1260 1280 1300
 ATATTATAG CAGAATCTCG AAACGTGTTA CTTTGCATCT AGCAGCTTGA GAATTGTAAC CTCTCCAGAC TAAGTAAGGA AGTGAAGAG AAAACCAAGC
 TATAATTATC GTCTTAGGAC TTTGACAAC GAAACGTAGA TCGTCGAAC CTTAACATTG GAGAGGTCTG ATTCATTCTC TCAGCTTCTG TTTTGGTTCG
 1320 1340 1360 1380 1400
 AGCTACGGTA TGGCTCCATT GATATGTTAT GCAGATAAAC CTATTTTCAT ATAGGCTATA GCTGTAAGAG ATCATCTATT TCATGTGTGT GGTTTTTTTT
 TCGATGCCAT ACCGAGGTA CTATACAACT CGTCTATTG GATAAAAGTA TATCCGATAT CGACATTCTC TAGTAGATA AGTACACACA CCAAAAAA
 1420 1440 1460 1480 1500
 TTTATGTTTT TTCAATGATG TGTGCATGCT ATTTTAGGT TTTAGAATCT ATTTTCATGGA AATTGAAGAT ATTTCAATTC ACGTGAAGT TCGTCAAGTT
 AAATACAAAA AAGTTACTAC ACACGTACGA TAAAAATCCA AAATCTTAGA TAAAGTACCT TTAACCTCTA TAAAGTAAAG TGCACATTCA AGCAGTTCAA
 1520 1540 1560 1580 1600
 GTGGCGTGTG TCTTGGAAAT TGATGTTTTG TTTGTAGATT TTAAGAGCTA CTCTTAAAT TTACAAGAGT TTTGTAATTT TCAATTATGG CCCATTATT
 CACCGCACAC AGAACCTTTA ACTACAAAAC AAACATCTAA AATTCTCGAT GAAGATTTTA AATGTTCTCA AAACATTAAA AGTTAATACC GGGTAATAAG
 1620 1640 1660 1680 1700
 TCATTAAATC ATTAATAAAA TTATATACAT TACTATCTAT ATCTAGCATA GGTAGTTTTT TTTTCTTTTT TCTTTGGTAG ACCTACTGAA CAAATATCTG
 AGTAATTAAG TAAATTTTTT AATATATGTA ATGATAGATA TAGATCGTAT CCATCAAAAA AAAAAGAAAA AGAAACCATC TGGATGACTT GTTTATAGAC
 1720 1740 1760 1780 1800
 ATATATCACT GACTGGATAA ATATCTATAG AGATATTTTT GATAGAAATG AGTGTTAATT TAACGTAAAA CAGGAAACTG AGAGGAGAGG ATCTTGATGG
 TATATAGTGA CTGACCTATT TATAGATATC TCTATAAAAA CTATCTTTAC TCACAATTA AATTGCAITTT GTCTTTTGAC TCTCCTCTCC TAGAACTACC
 1820 1840 1860 1880 1900
 ATTGAACCTA GAAGAGTTGC AGCGGCTGGA GAACTACTT GAATCCGGAC TTAGCCGTGT GTCTGAAAG AAGGTTTACT ACTATACATA AACTAATAGC
 TAACTTGAAT CTCTCAACG TCGCCGACCT CTCTGATGAA CTTAGGCGCT AATCGGCACA CAGACTTTTC TTCAAATGA TGATATGTAT TTGATTATCG
 1920 1940 1960 1980 2000
 ATGCATATTT TCCTTAACGT GGCATATAAA TAATAAGCTG TACATATATA AAAGTTTGAC TTTGTTGTTG TTATTGGTAA ATAGGGCGAG TGTGTGATGA
 TACGTATAAA AGGAATTGCA CGTATATTT ATTATTGAC ATGTATATAT TTTCAAACCTG AAACAACAAC AATAACCAT TATCCCGCTC ACACACTACT
 2020 2040 2060 2080 2100
 GCCAAATTTT CTCACTTGAG AAACGGGTTA GTAGTTAGTA CATACAATTC GTATAACTAA TGGATCATAA GCCTATCTAT AGCTAGTGAC TTTCTTAATA

CGGTTTAAAA GAGTGAACCT TTTGCCCAAT CATCAATCAT GTATGTTAAG CATATTGATT ACCTAGTATT CGGATAGATA TCGATCACTG AAAGAATTAT

2120 2140 2160 2180 2200

AGTGAACAG GGTACGGAT TGGTGGATGA GAATAAGAGA CTGAGGGATA AAGTACGGCT CTAAACCCCT ATAGATATCA TGGGAATAACC TTAATCTATT

TCACCTTTGTC CCTAGCCTTA ACCACCTACT CTTATTCTCT GACTCCCTAT TTCATGCCGA GATTTGGGAA TATCTATAGT ACCTTATTGG AATTAGATAA

2220 2240 2260 2280 2300

TTTTTATGTA TAAGAAAATA TGATGAGGGA ACGTATATTA TATATCGGCA GCTAGAGACG TTGSAAGGG CAAAACCTGAC GACGCTTAAA GAGGCTTTGG

AAAAATACAT ATTCTTTTAT ACTACTCCCT TGCATATAAT ATATAGCCGT CGATCTCTGC AACCTTTCCC GTTTTGACTG CTGCGAATTT CTCCGAAACC

2320 2340 2360 2380 2400

AGACAGAGTC GGTACCCACA AATGTGTCAA GCTACGACAG TGGAACTCCC CTTGAGGATG ACTCCGACAC TTCCCTGAAG CTTGGGTATA ATTTGTTTAA

TCTGTCTCAG CCACTGGTGT TTACACAGTT CGATGCTGTC ACCTTGAGGG GAACCTCTAC TGRGGCTGTG AAGGGACTTC GAACCCATAT TAAACAAATT

2420 2440 2460 2480

CTGAACATAT TTCAAACTTT TTGTTGACAT TTGTATGTG GATGTTTACT AACTGTTTGT TGGTTAGGCT TCCATCTTGG GAA

GACTTGTATA AAGTTTGAAA AACAACTGTA AAACATACAC CTACAAATGA TTGACAAACA ACCAATCCGA AGGTAGAACC CTT

SEQ ID NO:47

Arabidopsis AGL27 genomic sequence

CAACCAGCAG CACCAGCTGC AATCAAATCC TTTACGGTTC TTTGAATGTT TAGCGCATTT CCTCCACCGG TATCTTGAAA AGATCAAAAG AAACCTATGA

GTGGTTCGTC GTGGTCGACG TTAGTTTAGG AAATGCCAAG AACTTACAAA ATCGCGTAAA GGAGGTGGCC ATAGAACTTT TTTGATTAAT

-2881 -2861 -2841 -2821 -2801

AGAGAACTAT AACCAAGCAA ATCCACTATT TTCAAAAAGC TATGAAGAGA ACTATAAGCA AGCAAGCGAC TCTAACCAAG AAAGATTGAT ACTTTCAATC

TCTCTTGATA TTGGTTTCGTT TAGGTGATAA AAGTTTTCGT ATACTTCTCT TGATATTCGT TCGTTCGCTG AGATTGGTGT TTTCTAACTA TGAAGTTAG

-2781 -2761 -2741 -2721 -2701

TTTGGTAAAG AATCAACGAC TCAATGTTTT TAAATGTTTT TTTTCTTTT TTGGTTTATG TTAAGCTTCT TGCATTCTTT AATGATGTCT TTAATTACTT

AAACCAATTC TTAGTTGCTG AGTTACAAA ATTTACAAA AAAAGGAAAA AACCAAAATC AATTGGAAGA ACGTAAGAAA TTACTACAGA AATAATATGA

-2681 -2661 -2641 -2621 -2601

ATCAAAATTT TGCAACTTTA CCAGCATCTG CAATGATGGG TATATTAGGA GCTGACGCAC ACACCGACCT TGCCGTCGCA GCCATCTCCG GTGCTCTAAA

TAGTTTFAAA ACGTTGAAAT GGTCTGACAG GTTACTACCC ATATAATCTC CGACTGCGTG TGTGGCTGGA ACGGCAGCGT CGGTAGAGGC CACCAGATTT

-2581 -2561 -2541 -2521 -2501

ACGACGAAAG AACACAAATA AACCAAGAAC ATACAAAACA AAAATTACTA AAGAAAGAAA AAAAAAAGG TGCGCGACGT TAGCAAAACG AAATCGGGTT

TGCTGCTTTC TTGCTTTTAT TTTGCTTTCT TATGTTTGT TTTTAAATGAT TTCTTTCTTT TTTTCTTTCC ACCGCGTGCA ATCTGTTGGC TTTAGCCCAA

-2481 -2461 -2441 -2421 -2401

TTCCCGAGAG AGAAGCGGAT AAGGCGTAAC CGGATATATA ACCAGCGGAG AATCCGGTTT GCTGCACAAT AGCCGCGGAT AAGGCATCGT AGCATCCAGG

AAGGCTCCTC TCTTCGCTTA TTCCGCAATT GCCTATATTT TGGTCGCTCT TTAGGCCAAA CGACGTGTTA TCGGCGCTTA TTCCGTAGCA TCGTAGGTCC

-2381 -2361 -2341 -2321 -2301

CATAAGCACA ATGCCCTGTT CTTCATCATG CGCATGAAAA CGTGTGTTGA TTCTCGCTGT CGGATTCACC AATCTCGCGG CGCGTGGGTT CCGTCGGAAT

GTATTCTGTT TACGGAACAA GAAGTTAGTC CGCTACTTTT GCACAAACCT AAGAGCGACA GCCTAAGTGG TTAGAGCGGC GCGCACCCAA GGCAGCCTTA

-2281 -2261 -2241 -2221 -2201

GTGGTGAAG CTGTAAGGTT TAAGCTGCTA CAACAGAGTG AAGTTGTTTT GACAGCCATT AACATCGACA TTCTTCGAAG CCTCGAACAA GTTTTTCTCT

CAACCACTTC GACATTCCAA ATTGACGAT GTTGCTCACC TTCAACAAA CTGTGCGTAA TTGTAGCTGT AAGAAGCTTC GGAGCTTGTT CAAAAAAGAA

-2181 -2161 -2141 -2121 -2101

CTCTCTAATC GAGTTAGACT CTGACCCACA CGCTTGGGAT TTTAATAGAG AGCAGCTGGT TATTATATCT CGGTCTTATC TTATGGTAAC AGTATCTCAA

GAGAGATTAG CTCAATCTGA GACTGGGTGT GCGAACCCCTA AAATTATCTC TCGTGCACCA ATAATATAGA GCCAGAATAG AATACCAATT TCATAGAGTT

-2081 -2061 -2041 -2021 -2001

AGACTCAAAC CACAAGGTAT TGTGAAATG TTAGAGGCAA TCTAACAATA AATGTATAAT TTGGTTAGCT TAAGCTCATC ATAGAAATGG GCCTTTATGT

TCTGAGTTTG GTGTTCATA ACACCTTTAC AATCTCCGTT AGATTGTTAT TTACATATTA AACCAATCGA ATTGAGTAGT TATCTTTACC CGGAATACA

-1981 -1961 -1941 -1921 -1901

CACCAAACTT ATTTCAACAC ATAACACAAG AGCCACACAA ACAACGACTC CTTTCTCCAC CAGAACRAGC ACGACAAAGG CAAGAGAGTT GCAAAAGACC

GTGGTTTGGG TAAAGTGTGT TATTGTGTTT TCGGGTGTGT TGTGTCTGAG GAAAGAGGTG GTCTTGTTCG TGCTGTTTCC GTTCTCTCAA CGTTTTCTGG

-1881 -1861 -1841 -1821 -1801

TATAAGATGA TAACAATCGA AAAGATGTAA ATTTTGAGAA AAATCAAAAT AAACAAGAAA GATTTCAATT TTTTTCACCT TTTCTCCATT TCTACTTTGA

ATATTCTACT ATTGTAGCT TTTCTACATT TAAACTCTT TTTAGTTTAA TTTGTTCTTT CTAAAGTAAC AAAAAGTGAA AAAGAGGTAA AGATGAAACT

-1781 -1761 -1741 -1721 -1701

TTTTACATAC TCTATGGGCC AACCAATTTT CAACCTAATG CTTGATAAAA AATGATTTCG TTTTACTATC TCAACAAATT GGGCCTACAA CATCCAATTT

AAAAATGATG AGATACCCGG TTGGTTAAAG GTTGGAATTAC GAACATTTT TTAATAAGCC AAAATGATAG AGTTGTTTAA CCCGGATGTT GTAGGTTAAA

-1681 -1661 -1641 -1621 -1601

CATGTAGTGA CTTGTTTTCG CTTTTTTCAC ATCTCAACAA ATTGGGTCGT TTGTATTTAA GAAATTTGTA CAGCTTTTAA GACTGAATTT TACTTTATGG

GTACATCACT GAACAAAAAC GGAAAAAGTG TAGAGTTGTT TAACCCAGCA AACATAAATT CTTTAACAAT GTCGAAAAAT CTGACTTAAA ATGAATATCC

-1581 -1561 -1541 -1521 -1501

CITTTATGCT TCTTTTTCGG TTTTGATTAA GGGTGAATAT GTAAACTGTT GATACCATCT GATTTTTTTT ATTTTATTAT TTTCTTGTGT GCAACTATAC

GAAATACGAG AGAAAAAGGC AAAACTAATT CCCACTTATA CATTTGACAA CTATGTTAGA CTAAAAAAA TAAAAAATAA AAAGAACACA CGTTGATATG

-1481 -1461 -1441 -1421 -1401

CATCTGAATT CAATTGACAT TTTAGCCAAA TAAAAAGAT TGGTCCACTT GGATGGCTGT AAAAAAGTTT AGTGGAAAGT TTTATAGGGC TTGTGGCCAA

GTAGACTTAA GTTAACCTGA AAATCGGTTT ATTTTCTCTA ACCAGGTGAA CCTACCGACA TTTTTCCTAA TCACCTTCAT AAATATCCCG AACCAACGTT

-1381 -1361 -1341 -1321 -1301

TCTTCAACAA CGGCTATATA GTTGATCTTT TTAATAATTA ACTTACCGTT CGACTGTCTT CTCAACGATT TGACAAATTG COGTTAGATT AGTATTACTG

AGAAGTGGTT GCCGATATTA CAACTAGAAA AATTTTAATT TGAATGGCAA GCTGACAGAA GAGTTGCTAA ACTGTTAATC GGCRACTTAA TCATAATGAC

-1281 -1261 -1241 -1221 -1201

ATTTATTATT AACAAACCCA TTTCTTTTCT TATTTTGTAA TAAGCTAAAT CAGGCCAATA AAAGGGACAA GTAGAGATGG GCTATTTCTT TTTTCTCTCT

TAAATAATAA TTGTTTGGGT AAAGAAAAA ATAAAAAATT ATTGATTATA GTCCGCTTAT TTTCCCTGTT CATCTCTACC CGATAAAGAA AAAAGAGAAA

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

TTTTTTTTTC	-1181	GAGAAAAAGCC	-1161	GAGCTATCAT	-1141	TTAACCAGAA	-1121	GAAGCAAGCC	-1101
AAAAAAAAG	AATACATCAT	CTCTTTTCGG	AAATAAGAAT	CTCGATAGTA	AATGGTGGGT	AATGGTCTTT	CGACTCTTTA	CTTCGTCGG	CTTTGCTTAA
TGTAGTTTTG	-1081	GACGGTGAAG	-1061	GCATGTGAAT	-1041	GTCTTTTTGC	-1021	CGTTTAAAGG	-1001
ACATCAAAAC	CTGCCACTTT	TTATATCGGG	CCTTTAATGG	CGTACACTTA	TCTCAACTCT	CAGAAAAACG	GGGTTTATTA	GCAAATTCCT	TCATAACCGA
CGTTGGTTTA	-981	ATATTGGGCG	-961	GAAACGAGAT	-941	GTTTAAATCG	-921	CTGGTTCACC	-901
GCAACCAAA	TATAACCCGG	CTTTGCTCTA	ACCCTTCTTC	TTGTTACAGC	CAAAATTAGG	CAATCCGAGC	ACCCGACTAA	GACCAAGTGG	AAATATGCGA
AAGCGAACAA	-881	ACATTGAAAA	-861	TGGGGAAGCC	-841	AAATTAGTTA	-821	ACGCGCAAGC	-801
TTGCTTGTGT	TGTAACTTTT	ACCCCTTCGG	TTTAATCAAT	CCATCCCTAA	CTCAGTTTTG	AGACGTAGTA	TGAATGAGCC	TCACGACCTA	GATGCTGGAT
ACTCGATAAA	-781	GTAATGCTTA	-761	CTCTTGAGAG	-741	CGGAAGAAAG	-721	CAAAATCGCT	-701
TGAGCTATTT	CATTACCAAT	GAGAACCCTC	GCCTTCTTTC	CACAAAGATT	TTGATAAGGC	TTTCTAGTTG	GTGAAATGGT	CGGAGAGCCA	GCCTCTCGGT
TCATAGGAGC	-681	GGGAGGTGCA	-661	TATCTGAATA	-641	TCCAATGCA	-621	ACATTTTAAA	-601
AGTATCCTCG	CCCTCCACG	ATAGACTTAT	AGGGTACGT	TCAAGACAAG	ATGGATTTCG	AAAACAAAGA	AATTAACCAA	TGTAATAATT	TATACGAGAA
AGTTTTAGAT	-581	AATTAATGT	-561	TTTCAATACC	-541	CATCATCACT	-521	GGTGATAGTC	-501
TCAAAATCTA	TTATATTACA	AAAGTTATGG	TTAATAGAAT	CACGTATAGT	GGTCAAGTTA	CTAATCACTT	TTAATATTAA	CCACTATCAG	TTTGCTAATC
AAAATTATCG	-481	ATTTGAATTC	-461	AAAACCATTT	-441	TTTCTAGTTG	-421	CTTTAATATA	-401
TTTTAATAGC	TAAATTTTTA	TAAACTTAAG	TTTTGGTAAA	TAGTGAAGTT	TTGCAATGTA	GTTTGTATTA	TCCGATCAAT	GAAATTATAT	TAATCGAGTT
TAATAACTGA	-381	AATCCCTGAA	-361	CCCGATTTCAT	-341	AACCAATGAG	-321	TTTAAACCCC	-301
ATTATTGACT	TTAGGAACCT	AATTTGGCAAT	GGGCTAAGTA	AAGCACTACT	TTCCGATCAA	ATAAATAAAG	ATAAATAAAG	AAAATTGGGG	AGGTTTATTT
AAGAGAAAAC	-281	CTTAAAAAAC	-261	AATTTCTGTT	-241	CGTGCGGCGG	-221	AGCTTTTAAC	-201
TTCTCTTTTG	GAATTTTTTG	TTAAAGACAA	GCCACCCCTA	GATGATCGGA	CTCGGACCGG	TCTAACCGAC	TGGATTAAAA	AACGACAAGC	TTGCTGTTGG
TTAAAAATTT	-181	GCCTCTTAGT	-161	GGCTTCAAAA	-141	CTCTATCTCG	-121	AAAAAAAACG	-101
AATTTTTTAA	CGGAGAAATC	CGGAAGTTTT	GGCTTAGCAA	TCGCTTAATA	CTATTATTTT	GAGATAGAGC	AAATTGGTTT	TTTTTTTTGC	TCAACCTCTC
AAAAAAAAC	-81	CAAGAAAAAA	-61	GAATAAAAAA	-41	CGGAGATTAG	-21	CTTATCGGAG	-1
TTTTTTTTTG	GTTCCTTTTT	CTTATTTTTT	GTTTTTCGTA	TGAGCGTCTC	CGCTCTAATC	GATTAAATTA	GGGCATAACC	GAATAGCCCT	TAAACTTCGG
ATGGGAAGAA	-20	GAAAAATCGA	-40	GATCAAGCGA	-60	ACAAGTCACT	-80	TCTCATCGAC	-100
TACCCCTCTT	CTTTTTAGCT	CTAGTTCGCT	TAGCTCTTGT	TTTCGTCAGC	TGTTTCAGTA	AAGAGGTTTG	CTGCGTTACC	AGAGTAGCTG	TTTCGAGCTG
AACTTTTCGAT	-120	TCTCTGTGAA	-140	TCCTCCGTCG	-160	TCTATGACTC	-180	GACGAGTAAG	-200
TTGAAAGCTA	AGAGACACTT	AGGAGGCAGC	GACAACAGCA	CGATAGACGG	AGGCCTTTTG	AGATACTGAG	AAGGAGGCCA	CTGCTCAITC	TTCTATGAAA
CCTTTTCTGG	-220	GTCTCACTCG	-240	ATTTTGTGTC	-260	ATAGAAGCTT	-280	GCTTTTTTGT	-300
GGAAAGAGCC	CAGAGTGAGC	TAAAAACACG	AAAAAATGAA	TGTTTAATTA	CTTTCTCCAT	TATCTTCGAA	GTTTAGATCC	CGAAAAACTA	AGGTAGTTTA
CAACTGAGAT	-320	TTTCTCCTTG	-340	GAAGATAGCA	-360	ATTTAAGACT	-380	ATCGCCAAGA	-400
GTTGACTCTA	AAAGAGGAAC	AAAAGACATA	CTTCTATCGT	GATGCGTAAG	CTTTAACCTA	AAACATTTTG	TTGTGTAAGC	TAGCGGTTCT	ATACAAGAAC
ATGTTCTGTT	-420	CGTGTTTTTT	-440	TTTTTCTGTT	-460	TCTCTTTTTT	-480	TGTGACGAAA	-500
TACAAAGCAA	GCACAAAAAA	AAAAGCACAA	AAAAAATAAT	TCATTTTAAA	ATCATTTTAA	TACCTTCATT	ATGGAAGTAA	ACACTGCTTT	AAATTATAAC
CATGTTATTC	-520	AAGAAACTTT	-540	TCTACACGTG	-560	CTTTGTATTG	-580	CTGTTGTGTA	-600
GTACAATAAG	TTCTTTTAAA	AGATGTGCAC	CACATAAGCA	GTGATTTCGT	TTTAAGTAAT	CTAGTTCCAT	GATCAAGGTA	GACRACAGT	GAACCTTCGA
TGTTTTTTTC	-620	ATATAAGAAA	-640	CAATATGTTT	-660	TTTGGAATTT	-680	CTTCAAGTAG	-700
AGCAAAAAAG	TATATTCTTT	GTTATACAAA	TCTAACCAAG	AGATTGTGTA	TTTAACTACT	ATATTCAATA	TTGCAATGCA	GAAGTTCATC	AAAAACAATC
AGATTATTTG	-720	GGGTAGTAGG	-740	TAACATTAAT	-760	ATTTGACTTT	-780	AAAAATGATT	-800
TCTAATAAAC	CCCAATCACC	ATTGTAATTA	GCTTATAGAA	CGAATATCTT	TGGTTCAAA	ATTTGACTTT	ATGTTGATCC	TTTACATAAA	CATCTAGAAA
TCTTTTGTAA	-820	TTCTCTTTAA	-840	GGAATAAGGT	-860	TTTCTCTGAT	-880	TAAGTATGAG	-900
AGAAAAACAT	AAGAGAAATT	CCTTATTCCA	AATAGATCAA	TTATCTAGTT	GATTTTGATG	GTGCTGGGAT	AAGTTTCCAC	ATTGATACCT	GCCACACATT
CTTCATTACT	-920	TAACATAATT	-940	GATATCGATT	-960	TTTATGACAC	-980	ACATTATGTC	-1000
GAAGTAATGA	ATTGATTAA	CTATAGCTAA	AATTTGGAAA	TAATCGTAAT	TTGTTGTGTT	TAATCTGTG	GTATGTTCTA	TGTAATACAG	AATGACTCAC
ACTCTTTGTT	-1020	GCTCTCTAAG	-1040	ATGTTGTAGT	-1060	ATAACTGATT	-1080	CATATATATG	-1100
TGAGAAACAA	CGAGAGATTC	TACAACATCA	AACTCAAAGA	TTGCTAAAGA	AACTCAAAGT	TATTGACTAA	AATGACGATG	GTATATATAC	AGTCACCCGA
AGTAGGTTCA	-1120	TTAAGTAGAA	-1140	ATCGGTGCGC	-1160	TACAACCAAA	-1180	CTTTAATAGT	-1200
TCATCCAAAT	AATTCATCTT	TAGCCAGCGG	TTAAATGAT	AACTCTCTCT	TGGTGTCTG	ATGTTGGTTT	ACAAGTTACT	GAAATTATCA	GAAGACAATA
TTGTCGTGGA	-1220	TATTTTAAAC	-1240	CCCATGAAC	-1260	CTTTTAGAAT	-1280	CGACTAAAGG	-1300
AACAGCACCT	ATAAAAAATT	GGGTACTTGA	AAACATAGAT	TTTGTATCTA	GAATAATCTC	ATCCACTTCT	ACTTTGAATG	TGAGTTTCTT	ACTCAAAAAA
	1320		1340		1360		1380		1400

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

TTTTCTAATA GACCTAAGAT AAAATCATCA ATGGATAAGT AGGAAATGGA AAGGTAACCT TTGTCAGTAT GTGTATATAT ACAGCTCCTT CTCATTTTCTT
AAAAAGATTAT CTGGATTCTA TTTTAGTAGT TACCTATTCA TCCTTTTACCT TTCCATTGAG AACAGTCATA CACATATATA TGTCCGAGGAA GAGTAAAGGA

1420 1440 1460 1480 1500
TGATGTTGAC TCCATAAATG CTTGATCATG AAAGCAAATT TGTAAATTT GTAACCAACA AAATGCACAG ACTATAGACG AAGTATTAGG AACCGTATCT
ACTACAACCTG AGGTATTATC GAAGTAGTAC TTTTCGTTTAA ACAATTTTAA CATTGGTTGT TTTACGTGTC TGATATCTGC TTTCAATATCC TTGGCATAGA

1520 1540 1560 1580 1600
ATCTGCTCC ATTTTACAAT AGTCAAGCTC TAGTTGTAGC TAGTTTCTTT ATTTAGTTCT TATACCTTAA CAAAGTGGCA CTATGCAAG TGTTTTGTAGT
TAGACAGAGG TAAATGTTA TCAGTTCGAG ATCAACATCG ATCAAGAGAA TAAATCAAGA ATATGGAATT GTTTCACCGT GATACGTTTC ACAAAATCA

1620 1640 1660 1680 1700
TGAGATTAGT CGTCTTATGC GTCTTACTAA TTGTTCACTT TTTCTTCTTT TTGTGATTGA TGTAAAATTA CTAAGTCACA ACTTGAGATG TTACTAAAAA
ACTCTAATCA GCAGAAATAC CAGAATGATT AACAAAGTAAA AACACTAACT ACATTTTAAAT GATTTCAGTGT TGAACCTTAC AATGATTTT

1720 1740 1760 1780 1800
GATAAGAACG TGTAAATACT GAAGTGAATT TGAAGCCAGT CTCTATTTCAT ATCATAGCAT TAATAGATCA TGGACAACAC ATATATAGGA TTAGAGCTGT
CTATTCTTGC ACATTATTGA CTTCACTTAA ACTTCGGTCA GAGATAAGTA TAGTATCTAGT ATTATCTAGT ACCTGTTGTG TATATATCTCT AATCTCGACA

1820 1840 1860 1880 1900
CATGACCTTC CGGAAATGC TAAATCAGTT TCTTGGTTTA TCCTTTTTCG AGTATCATGA TATCATTTAG CCAAAGGTTT TTGTTTTCAG TATTCGGATT
GTACTGGAAG GGCCTTTACG ATTTAGTCAA AGAACCAAAAT AGGAAAAACC TCATAGTACT ATAGTAAATC GGTTTCCAAA AACCAAAGTC ATAAGGCTAA

1920 1940 1960 1980 2000
CGTTTGACGT TATGTGTGAA AGCGTCAATA ACTAAACTT GGATTGACTA GTCAAAATAT AAAGTGAATT CATTGAATTC TTGAAAATTT TCCCTTAAAA
GCAAACTGCA ATACACACTT TCGCAGTTAT TGATTTTGAA CCTAAGTAT CAGTTTATATA TTTGACTAAC GTAACTTAAG AACTTTTAAA AGGGAATTTT

2020 2040 2060 2080 2100
TGAACATGAA TTTTCAAG ATTTTGTCTT TTGGAGGAT GTGATTTATA ATCTATACAA TCATACATTT TGCATGATAT TAGTTTTTTG AAGAACCATA
ACTTGTACTT AAAGTAGTTC TAAAAACAGAA AACCTTCTTA CACTAAATAT TAGATATGTT AGTATGTAAA ACGTACTATA ATCAAAAAAC TTCTTGGTTT

2120 2140 2160 2180 2200
AATAGAGCTT CTTTATAAAG CTGATTTAGC CTTGATAAGA AAAAGAAGGT AGATAATCGA ACTCATGGGG ATGAGTTTAA AATGTGTGCA CTTAGTTTCT
TTATCTCGAA GAAATATTTT GACTAAATCG GAACATTTCT TTTTCTTCCA TCTATTAGCT TGAGTACCCC TACTCAATTT TTACACACGT GAATCAAGA

2220 2240 2260 2280 2300
AAAACCTTTT GAAGTCGAAA CAATGACAAT ATTGCGTCCG AAGTTGATAT ATAACAGCAT CTTAAAGTTG AAATTGTAAA TTCAGATTTT AATTTTAGAG
TTTTGGAAAA CTTTCAGCTTT GTTACTGTTA TAACGACGCG TTCAACTATA TATTGTCTTA GAATTTCAAC TTTAACATTT AAGTCTAAAA TTAATATCTC

2320 2340 2360 2380 2400
CACCAGATGA TCAGAGTTTC AGATTTACAT TTGAAGTATA AAACATTTTG AACACATATA TCTAAAGCAG TAACCTTCAA AATAGGGTAA CTAATAGTAA
GTGGTCTACT AGTCTCAAG TCTAAATGTA AACTTTCATAT TTTGTAAAAC TTGTGTATAT AGATTTTCGTC AITGAAAGTTT TTATCCCAT TATTATCAT

2420 2440 2460 2480 2500
CTTACATTGT TTTTATTAAT GCTTTTATAC TTACTATCAT TTTTATATAT AGATGCCTGG TTAAGTAAAG ATGATTATCA AAAAGCTGTT GTTAGTAAAC
GAATGTAAAC AAAAAAATTA CGAAAAATAT AATGATAGTA AAAATATATA TCTACGGAAC AATTCATTTT TACTAATAGT TTTTGACAAC CAATCATGT

2520 2540 2560 2580 2600
GAAATGCTTG CAAATGTAA ATATTATATA AGCTTTCTTT CACTTTGGTG CATTCTCTCT AAATAATGGC CTCTATTGAT CGAGTATCTG AITCTTAGTT
CTTTAAACAC GTTTACATTG TATAATATAT TCGAAAGAAA GTGAAGCCAC GTAAAGAGAGA TTTATTACCG GAGATAACTA CGTCATAGAC TAAGATACAA

2620 2640 2660 2680 2700
TTGAAATGGT TTTTGCATAA ATTTATGTTT TAATGCAATT TTGTTTATC TCCAGCATTT CCAAGTATCAT TGATCGTTAT GAAATACAAC ATGCTGATGA
AACTTTTACCA AAAACGTATT TAATAACAAG ATTACGTAAA AACAAAATAG AGGTGTTAAA GGTTCATGTA ACTAGCAATA CTTTATGTTG TACGACTACT

2720 2740 2760 2780 2800
ACTTAGAGCC TTAGTAAGTA ATTAGCTAAG AACGTCATTC TAATATTTCT CTGGATGCGG TTTTTGTGTG TATGAAGGAT AGAAGCGCTG TTCAAGCCGG
TGAATCTCGG AATCATTCAT TAATCGATT TTGCAAGTAA ATTATAAGAA GACCTACGCC AAAAACCCACA ATACTTCCCTA TCTTCGCGAC AAGTTCGGCC

2820 2840 2860 2880 2900
AGAAACCTCA ATGTTTGAAT CTGTAACAC CGAATTTAAT TCTCTAGAGT TACAGTTATT GTGTCTACTG GAAATACAA GAACTTCACA ATCTTTCTGA
TCTTTGGAGT TACAAAACCT GAGCATTGTG GCTTGAATTA AGAGATCTCA ATGTCAATA CACAGATGAC CTTTATGTT CTGGAAGTGT TAGAAGAGCT

2920 2940 2960 2980 3000
CCATTCCTTT TCTTCAATGG CAGGATCTTG AAGAAAAAAT TCAGAAATTAT CTTCCACACA AGGAGTTACT AGAAACAGTC CAAAGGTTAG CAGTACGACA
GGTAAGGAAA AGAAGTACAC GTCCTAGAAC TTTCTTTTAA AGTCTTAATA GAAGGTGTGT TCTCAATGA TCTTTGTGAG GTTCCAAATC GTCATGCTGT

3020 3040 3060 3080 3100
CATTTTTCTC CCCTCTCTCT CTGATAAAAA AAATGTTTTT TTTCTTTTGT CTACTTGTGA ATACAGCAAG CTTGAAGAAG CAAATGTGCA TAATGTAAAT
GTAAAAAGAG GGGAGAAAGA GACTATTTTT TTTACAAAAA AAAGAAAAACA GATGAACACT TATGTCGTTT GAACTTCTTG GTTTACAGCT ATTACATTCA

3120 3140 3160 3180 3200
GTAGATTCTC TAATTTCTCT CGAGGAACAA CTTGAGACTG CTCTGTCCGT AAGTAGAGCT AGGAAGGTAT ATGTGCTGCT ACTAAGTGAT TCAACCAATT
CATCTAAGAG ATTAAGAGA CCTCCTTGTT GAACTCTGAC TTTATCTGTA TTTATCTGTA TCCCTCCATA TACACGAGCA TGATTTCACTA AGTTGGTTAA

3220 3240 3260 3280 3300
ACTCCACAAA ACCTTCTTTT TAGTTAGTTA TCCTAGAACAA ATCTTTTGAC ATAAATCTTA ATGTCTTGTT ATAGGCAGAA CTGATGATGG AGTATATCGA
TGAGGTGTTT TGGAAAGAAA ATCAATCAAT AGGATCTTGT TAGAAAACTG TATTTAGAAAT TACAGAACAA TATCCGCTCT GACTACTACC TCATATAGCT

3320 3340 3360 3380 3400
GTCCCTTAAA GAAAGGTTA GTGCTTTGGT TTTTATTTTC GATAAAGGCC ATATTCTAGG CTATGATGAT TCTTGAATTC TATTAACCTG CTGAGTCTAC
CAGGGAATTT CTTTCCCAAT CACGAAACCA AAAATAAAAG CTATTTCCGG TATAAGATCC GATACTACTA AGAATTTAAG ATAATTGGAC GACTCAGATG

3420 3440 3460 3480 3500
AGATTACTAT ATATATATAT ATATATCTTT TGGTCTTGTC TTAGTCTCTG ATTTAGTATT GGTTCATTC AGGTGAAACC CTAATGAGAA TTAATAAAC
TCTAATGATA TATATATATA TATATAGAAA ACCAGAACAG AATCAAGGAC TAAATCATAA CCGAAGTAAG TCCACTTTGG GATTACTCTT AATTTTGTG

3520 3540 3560 3580 3600
AAGCAGTTTT AAACCTCTGA TCBAATCCAA CCTTCCCTC ATAAAGTGTC GAATTTGGAT GAGGATGATT TATGTTTCCA GAAGGAAACA TGTTTGGAAG
TTGCTCAAAA TTTGAGAACT AGTTTAGGTT GGAAGGGGAG TATTTACAGC CTTAAACCTA CTCCTACTAA ATACAAAGCT CTTCCTTTGT ACAAACTTTT

3620 3640 3660 3680 3700
TAGCTATAGA AGTTGTTAGA AACTAATGAC CTTATGATCT TTTCCAAACA GGAGAAATTT CTGAGAGAAG AGAACCAGGT TCTGGCTAGC CAGGTAAACA
ATCGATATCT TCAACAATCT TTGATTACTG GAATACTAGA AAAGGTTTGT CCTCTTAAAC GACTCTCTTC TCTTGGTCCA AGACCGATCG GTCCATTGTT

3720 3740 3760 3780 3800
TGACCACAAT ATCTTCTGCT CTTGAAGCTA ATTAATCACT TTATACGTC CCGTTATAGA GAGATACACA TATACACGTA CATGAAACT AAAAGTTTGA
ACTGGTGTTA TAGAAGACGA GAACCTCGAT TAATTAGTGA AATATGCAGG GGCAATATCT CTCTATGTTG ATATGTGCAT GTACTTTTGA TTTTCAACTT

3820 3840 3860 3880 3900
GGACTTTGAT GGATACAGA CAATTATAGT GAAACCTTAA ATATGTGATA AGTGATAACA AAATGCTTTT AAAATCTATC TTTCTTGTTA ATTTAGTAGC
CTGAAACTA CCTATGATCT GTTAATATCA CTTTGGGATT TATACACTAT TCATATTGTT TTTACGAAAA TTTTAGATAG AAAGAACAA TAAATCATCG

3920 3940 3960 3980 4000
 TGTGAGAGAA GAAAGGTATG TCTCACCAGT GAAAGATACT CAAAACCCGG TATTTTAAAT TTGTGAAATT TGCAAAATAA AAAAATGCTT TCTACAAGAT
 ACAGTCTCTT CTTTCCATAC AGAGTGGCTA CTTTCTATGA GTTTTGGGCC ATAAAAATTA AACACTTTAA ACGTTTATTT TTTTACGAA AGATGTTCTA
 4020 4040 4060 4080 4100
 AGATTAAATTT CTTGCAATGT TTAGTAGCTG TAGAAAAAAA AGAAATGTAA GAAAGTTTCT TACAGATGGG AAAGAATACG TTGCTGGCAA CAGATGATGA
 TCTAATTAAG GAACGTTACA AATCATCGAC ATCTTTTCTT TCTTTACATT CTTTCAAAGA ATGTCTACCC TTTCTTATGC AACGACCGTT GTCTACTACT
 4120 4140 4160 4180 4200
 GAGAGGAATG TTTCCGGGAA GTAGCTCCGG CAACAAAATA CCGGAGACTC TCCCGCTGCT CAATTAGCCA CCATCATCAA CGGCTGAGTT TTCACCTTAA
 CTCCTCTTAC AAAGGCCCTT CATCGAGGCC GTTGTTTTAT GGCCTCTGAG AGGGCGACGA GTTAATCGGT GGTAGTAGTT GCCGACTCAA AAGTGAATT

SEQ ID NO:48 and SEQ ID NO:49

Alternatively splice Arabidopsis AGL27 cDNA and resulting Alternate Arabidopsis AGL27 amino acid sequence

20 40 60 80 100
 ATGGAAGAA GAAAAATCGA GATCAAGCGA ATCGAGAACA AAAGCAGTCG ACAAGTCACT TTCTCCAAAC GACGCAATGG TCTCATCGAC AAAGCTCGAC
 TACCCTTCTT CTTTITAGCT CTAGTTCGCT TAGCTCTTGT TTTCTGTCAGC TGTTCACTGA AAGAGGTTTG CTGCGTTACC AGAGTAGCTG TTTGAGCTG
 M G R R K I E I K R I E N K S S R Q V T F S K R R N G L I D K A R>
 120 140 160 180 200
 AACTTTCGAT TCTCTGTGAA TCCTCCGTCG CTGTTGTCTG CGTATCTGCC TCCGGAAAAAC TCTATGACTC TTCTCCCGGT GACGAGATAG AAGCGCTGTT
 TTGAAAGCTA AGAGACACTT AGGAGGCAGC GACAACAGCA GCATAGACGG AGGCCTTTTG AGATACTGAG AAGGAGGCCA CTGCTCTATC TTCGCGACAA
 Q L S I L C E S S V A V V V V S A S G K L Y D S S S G D E I E A L F>
 220 240 260 280 300
 CAAGCCGGAG AAACCTCAAT GTTTTGAACG CGATCTTGAA GAAAAAATTC AGAATTATCT TCCACACAAG GAGTTACTAG AAACAGTCCA AAGCAAGCTT
 GTTCGGCCTC TTTGGAGTTA CAAAACCTGA GCTAGAACTT CTTTTTAAAG TCTTAATAGA AGGTGTGTTC CTCAATGATC TTTGTCAGGT TTCGTTTCGAA
 K P E K P Q C F E L D L E E K I Q N Y L P H K E L L E T V Q S K L>
 320 340 360 380 400
 GAAGAACAA ATGTCGATAA TGTAAGTGTA GATTCTCTAA TTTCTCTGGA GGAACAACCT GAGACTGCTC TGTCGGTAAG TAGAGCTAGG AAGGCAGAAC
 CTTCTTGGTT TACAGCTATT ACATTACAT CTAAGAGATT AAAGAGACCT CCTTGTGTA CTTCTGACGAG ACAGGCATTC ATCTCGATCC TTCGCTCTTG
 E E P N V D N V S V D S L I S L E E Q L E T A L S V S R A R K A E>
 420 440 460 480 500
 TGATGATGGA GTATATCGAG TCCCTTAAAG AAAAGGAGAA ATTGCTGAGA GAAGAGAACC AGTTTCTGGC TAGCCAGATG GGAAAGAATA CGTTGCTGGC
 ACTACTACCT CATATAGCTC AGGGAATTTT TTTTCTCTT TAACGACTCT CTTCTCTTGG TCCAGACCG ATCGGTCTAC CTTTCTTAT GCAACGACCG
 L M M E Y I E S L K E K E K L L R E E N Q V L A S Q M G K N T L L A>
 520 540 560
 AACAGATGAT GAGAGAGGAA TGTTTCCGGG AAGTAGCTCC GGCAACAAAA TACCGGAGAC TCTCCCGCTG CTCAATTAG
 TTGCTCTACT CTCTCTCTT ACAAGGCCCT TTATCGAGG CCGTTGTTTT ATGGCCTCTG AGAGGGCGAC GAGTTAATC
 T D D E R G M F P G S S S G N K I P E T L P L L N *>